
WATERMAN

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm
on: Fri Sep 17 21:10:05 1999; MasPar time 39.18 Seconds
Tabular output not generated. 1020.796 Million cell updates/sec

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pep
Perfect Score: 9849
Sequence: 1 MSGLGENLDPLASDSRKRL.....MNNPMPMSGMPGDPQKYC 1415

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
l:swissprot

Statistics: Mean 56.586; Variance 121.726; scale 0.465

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	232	2.4	805	1	AHR_MOUSE	AH RECEPTOR PRECURSOR	9.74e-17
2	232	2.4	848	1	AHR_HUMAN	AH RECEPTOR (ARYL HYDR	9.74e-17
3	239	2.4	853	1	AHR_RAT	AH RECEPTOR (ARYL HYDR	7.91e-18
4	225	2.3	339	1	TF2D_HUMAN	TRANSCRIPTION INITIATI	1.17e-15
5	224	2.3	642	1	ARNT_DROME	ARYL HYDROCARBON RECEP	1.67e-15
6	230	2.3	644	1	BTB_DROME	TRANSCRIPTION FACTOR B	1.99e-16
7	227	2.3	758	1	Y438_YEAST	HYPOTHETICAL 85.0 KD P	5.77e-16
8	226	2.3	905	1	SNF5_YEAST	TRANSCRIPTION REGULATO	8.23e-16
9	229	2.3	1794	1	YAV1_SCHPO	HYPOTHETICAL 193.3 KD	2.84e-16
10	219	2.2	262	1	GBA1_WHEAT	ALPHA/BETA-GLIADIN PRE	9.70e-15
11	215	2.2	313	1	GBA7_WHEAT	ALPHA/BETA-GLIADIN PRE	3.93e-14
12	221	2.2	609	1	OPA_DROME	PAIR-RULE PROTEIN ODD-	4.81e-15
13	213	2.2	2124	1	Y192_HUMAN	HYPOTHETICAL PROTEIN K	7.89e-14
14	217	2.2	2175	1	HMCU_DROME	HOMEBOX PROTEIN CUT.	1.96e-14
15	210	2.1	360	1	MDL1_HUMAN	MACHADO-JOSEPH DISEASE	2.23e-13
16	205	2.1	467	1	INVO_MOUSE	INVOLUCRIN.	1.25e-13
17	209	2.1	594	1	NPAL_MOUSE	NEURONAL PAS DOMAIN PR	3.15e-13
18	205	2.1	648	1	KAPC_DICDI	CAMP-DEPENDENT PROTEIN	1.25e-12
19	204	2.1	708	1	GFE_DICDI	G-BOX BINDING FACTOR (1.76e-12
20	202	2.1	870	1	PAS1_HUMAN	ENDOTHELIAL PAS DOMAIN	3.49e-12
21	211	2.1	874	1	PAS1_MOUSE	ENDOTHELIAL PAS DOMAIN	1.58e-12
22	204	2.1	966	1	SSN6_YEAST	GLUCOSE REPRESSION MED	1.76e-12
23	204	2.1	1081	1	GALY_YEAST	TRANSCRIPTION REGULAO	1.76e-12

RESULT ID	AHR_MOUSE	STANDARD;	PRT;	805 AA.
AC	P30561;			
DT	01-APR-1993	(REL. 25, CREATED)		
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998	(REL. 37, LAST ANNOTATION UPDATE)		
DE	AH RECEPTOR PRECURSOR (ARYL HYDROCARBON RECEPTOR).			
GN	AHR.			
OS	MUS MUSCULUS (MOUSE).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;			
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 10-25.			
RX	MEDLINE; 92231934.			
RA	EMA M., SOGAWA K., WATANABE N., CHUOH Y., MATSUSHITA N., GOTOH O., FUNAE Y., FUJII-KURIYAMA Y.;			
RT	"CDNA cloning and structure of mouse putative Ah receptor.";			
RL	BIOCHEM. BIOPHYS. RES. COMMUN. 184:246-253(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN-C57BL;			
RX	MEDLINE; 92390411.			
RA	BURBACH K.M., POLAND A., BRADFIELD C.A.;			
RT	"Cloning of the Ah-receptor cDNA reveals a distinctive			
RL	ligand-activated transcription factor.";			
RN	[3]			
RP	PROC. NATL. ACAD. SCI. U.S.A. 89:8185-8189(1992).			
CC	SEQUENCE FROM N.A.			
RC	STRAIN-DBA/2J; TISSUE=LIVER;			
RA	EMA M., OHE N., SUZUKI M., MIMURA J., SOGAWA K., IKAWA S., FUJII-KURIYAMA Y.;			
RL	SUBMITTED (SEP-1994) TO EMBL/GENBANK/DBJ DATA BANKS.			
CC	-1- FUNCTION: THE AH RECEPTOR ACTIVATED IN ASSOCIATION WITH A LIGAND			
CC	IS CONSIDERED TO TRANSLocate FROM CYTOPLASM TO NUCLEUS AND			
CC	ENHANCE THE TRANSCRIPTION OF THE GENES BY BINDING TO THE XRE			
CC	SEQUENCE IN THEIR PROMOTER REGIONS. IT MEDIATES BIOCHEMICAL AND			
CC	TOXIC EFFECTS OF HALOGENATED AROMATIC HYDROCARBONS.			
CC	-1- SUBCELLULAR LOCATION: FIRST CYTOPLASMIC, UPON BINDING WITH LIGAND			
CC	AND INTERACTION WITH A HSP90, IT TRANSLOCATES TO THE NUCLEUS.			
CC	-1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER			
CC	BLH PROTEIN. IN THE NUCLEUS, HETERODIMER OF AHR AND ARNT.			
CC	-1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.			
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF			
CC	TRANSCRIPTION FACTORS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; D38417; G1215804; -
DR EMBL; M94623; G192101; -
DR EMBL; D38416; -; NOT_ANNOTATED_CDS.
DR PIR; JQ1485; JQ1485.
DR PIR; A46266; A46266.
DR MGI; MGI:105043; AHR.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR PFAM; PF00785; PAC; 1.
DR PFAM; PF00989; PAC; 1.
DR TRANSFAC; T00018; -.
DR TRANSFAC; T00194; -.
KW TRANSCRIPTION REGULATION; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
FT PROPEP 1 9

CHAIN 10 805 AHR RECEPTOR.
DNA_BIND 27 39 BASIC DOMAIN.
DOMAIN 40 80 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
REPEAT 111 177 PAS-1.
FT REPEAT 269 336 PAS-2.
FT DOMAIN 342 380 PAC MOTIF.
FT DOMAIN 594 648 GLN-RICH.
FT VARIANT 324 324 I -> M (IN DBA/2J).
FT VARIANT 351 351 H -> N (IN DBA/2J).
FT VARIANT 375 375 A -> V (IN DBA/2J).
FT VARIANT 471 471 L -> P (IN DBA/2J).
FT VARIANT 533 533 S -> N (IN DBA/2J).
FT VARIANT 589 589 L -> M (IN DBA/2J).
FT VARIANT 805 805 S -> SRGIENETYS (IN DBA/2J).
FT CONFLICT 74 74 S -> T (IN REF. 2).
SQ SEQUENCE 805 AA; 90337 MW; 1998DIAA CRC32;

Query Match 2.4%; Score 232; DB 1; Length 805;
Best Local Similarity 30.3%; Pred. No. 9.74e-17;
Matches 46; Conservative 46; Mismatches 52; Indels 8; Gaps 6;

Db 26 PARGIKSNPSKRRDRRLNTELDRLASLLP--FPQ-DVIN-KLDKLSVLRLSVYLRKSF 81

QY 25 PGQGLTCSGKRRRQESKYIEELAEISANLSDIDNFNVKPKCAILKETVRIQRIKE 84

Db 82 FDVALKSTPADRNGGQDCQACQ--IRWDQLQEGEFLQALNGFLVVTADALVFFASSTI 140

QY 85 QGKTI-SNDDVOKADVSTGQGVDRKSL--GPLLLQALDGLFVFNREANIVFVSENV 141

141 QDYLGFQQSDVIHQSVYELIHTEDRAEFQRL 172

142 TQYLYQKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 2

ID AHR_HUMAN STANDARD; PRT; 848 AA.

AC P35869;

DT 01-JUN-1994 (REL. 29, CREATED)

DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE AHR RECEPTOR (ARYL HYDROCARBON RECEPTOR).

GN AHR.

OS HOMO SAPIENS (HUMAN).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LIVER;

RX MEDLINE; 93347997.

RA ITOH S., KAWATAKI T.;

RT "Human Ahr receptor cDNA: analysis for highly conserved sequences.";

RL NUCLEIC ACIDS RES. 21:3578-3578(1993).

RN [2]

RP SEQUENCE FROM N.A.

RX

RA

RT

RL

RN

RP

RA

RL

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

MEDLINE; 94067047.

DOLANICK K.M., SCHMIDT J.V., CARVER L.A., SWANSON H.I., BRADFIELD C.A.;

"Cloning and expression of a human Ahr receptor cDNA.";

MOL. PHARMACOL. 44:911-917(1993).

[3]

SEQUENCE FROM N.A.

ANTONIOU B., SMITH A., ELLIOTT G., KRAMER J.;

SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.

FUNCTION: THE AHR RECEPTOR ACTIVATED IN ASSOCIATION WITH A LIGAND

IS CONSIDERED TO TRANSLOCATE FROM CYTOPLASM TO NUCLEUS AND

ENHANCE THE TRANSCRIPTION OF THE GENES BY BINDING TO THE XRE

SEQUENCE IN THEIR PROMOTER REGIONS. IT MEDIATES BIOCHEMICAL AND

TOXIC EFFECTS OF HALOGENATED AROMATIC HYDROCARBONS.

CELLULAR LOCATION: FIRST CYTOPLASMIC. UPON BINDING WITH LIGAND

AND INTERACTION WITH A HSP90, IT TRANSLOCATES TO THE NUCLEUS.

SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER

BHLH PROTEIN. IN THE NUCLEUS, HETERODIMER OF AHR AND ARNT.

FUNCTION: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF

TRANSCRIPTION FACTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D16354; G533324; -

EMBL; L19872; G416142; -

EMBL; AC003075; G2588638; -

MIM; 600253; -

PROSITE; PS00038; HELIX_LOOP_HELIX; 1.

PFAM; PF00785; PAC; 1.

PFAM; PF00989; PAC; 1.

TRANSFAC; T01795; -

TRANSCRIPTION REGULATION; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.

DNA_BIND 28 40 BASIC DOMAIN.

DOMAIN 41 81 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).

REPEAT 113 179 PAS-1.

REPEAT 275 342 PAS-2.

DOMAIN 348 386 PAC MOTIF.

DOMAIN 600 640 GLN-RICH.

CONFLICT 807 848 LNEYPAELNNINNTTTHLOPLHHPSEARFPDLTSSGF

L -> FK (IN REF. 1).

SEQUENCE 848 AA; 96147 MW; 51BB7808 CRC32;

Query Match 2.4%; Score 232; DB 1; Length 848;

Best Local Similarity 27.5%; Pred. No. 9.74e-17;

Matches 42; Conservative 52; Mismatches 50; Indels 9; Gaps 7;

Db 27 PARGIKSNPSKRRDRRLNTELDRLASLLP--FPQ-DVIN-KLDKLSVLRLSVYLRKSF 82

QY 25 PGQGLTCSGKRRRQESKYIEELAEISANLSDIDNFNVKPKCAILKETVRIQRIK- 83

83 FDVALKSSPTERNQDNCRAANFREGNLQOE-GEFLQALNGFLVVTADALVFFASST 141

84 -EKG-KTISNDDVOKADVSTG-QGVDRKSL--GPLLLQALDGLFVFNREANIVFVSENV 140

142 IODYLGFGQSDVIHQSVYELIHTEDRAEFQRL 174

141 TQYLYQKQEDLVNTSVYNILHEEDRKDFLKNL 173

RESULT 3

ID AHR_RAT STANDARD; PRT; 853 AA.

AC P41738;

DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)

DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)

DE AHR RECEPTOR (ARYL HYDROCARBON RECEPTOR).

GN AHR.

85	QGKTI-SNDDVQKADVSSTGGQVDDKDSL--GPLLLQALDQFLFVVYVNRANIVFVSEN	141
141	QDYLFGQQSDVIHQSVYELIHTEDRAEFQRL	172
142	TOYLOVQEDLVNTSVYNILHEEDRKDFLKNL	173
4		
RESULT		
ID	TF2D_HUMAN	STANDARD;
AC	P20226;	PRT; 339 AA.
DT	01-FEB-1991 (REL. 17, CREATED)	
DT	01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)	
DE	TRANSCRIPTION INITIATION FACTOR TF2ID (TATA-BOX FACTOR) (TATA	
DE	SEQUENCE-BINDING PROTEIN) (TBP).	
OS	TBP OR TF2ID OR TF2D.	
OS	HOMO SAPIENS (HUMAN).	
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;	
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.	
CC	[1]	
CC	SEQUENCE FROM N.A.	
CC	MEDLINE; 90302010.	
CC	KAO C.C., LIEBERMAN P.M., SCHMIDT M.C., ZHOU Q., PEI R., BERK A.J.;	
CC	"Cloning of a transcriptionally active human TATA binding factor.";	
CC	SCIENCE 248:1646-1649(1990).	
CC	[2]	
CC	SEQUENCE FROM N.A.	
CC	MEDLINE; 90326195.	
CC	HOFFMANN A., SINN E., YAMAMOTO T., WANG J., ROY A., HORIKOSHI M.,	
CC	ROEDER R.G.;	
CC	"Highly conserved core domain and unique N terminus with presumptive	
CC	regulatory motifs in a human TATA factor (TF2ID).";	
CC	NATURE 346:387-390(1990).	
CC	[3]	
CC	DOMAINS.	
CC	MEDLINE; 90302006.	
CC	PETERSON M.G., TANESE N., PUGH B.F., TJIAN R.;	
CC	"Functional domains and upstream activation properties of cloned	
CC	human TATA binding protein.";	
CC	SCIENCE 248:1625-1630(1990).	
CC	[4]	
CC	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 159-337.	
CC	MEDLINE; 96209823.	
CC	NIKOLOV D.B., CHEN H., HALAY E.D., HOFFMANN A., ROEDER R.G.,	
CC	BURLEY S.K.;	
CC	"Crystal structure of a human TATA box-binding protein/TATA element	
CC	complex.";	
CC	PROC. NATL. ACAD. SCI. U.S.A. 93:4862-4867(1996).	
CC	[5]	
CC	X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS) OF 159-339.	
CC	MEDLINE; 96346176.	
CC	JUO Z.S., CHIU T.K., LETBERMAN P.M., BAIKALOV I., BERK A.J.,	
CC	DICKERSON R.E.;	
CC	"How proteins recognize the TATA box.";	
CC	J. MOL. BIOL. 261:239-254(1996).	
CC	!- FUNCTION: GENERAL FACTOR THAT PLAYS A MAJOR ROLE IN THE ACTIVATION	
CC	OF EUKARYOTIC GENES TRANSCRIBED BY RNA POLYMERASE II. TF2ID BINDS	
CC	SPECIFICALLY TO THE TATA BOX PROMOTER ELEMENT WHICH LIES CLOSE TO	
CC	THE POSITION OF TRANSCRIPTION INITIATION.	
CC	!- SUBUNIT: BINDS DNA AS A MONOMER.	
CC	!- SUBCELLULAR LOCATION: NUCLEAR.	
CC	!- SIMILARITY: THE C-TERMINAL 180 RESIDUES ARE EXTREMELY WELL	
CC	CONSERVED IN ALL EUKARYOTIC TF2ID	
CC	!- SIMILARITY: WEAK, WITH BACTERIAL POLYMERASE SIGMA-FACTORS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	

```
DR EMBL; M55654; G339492; -.
DR EMBL; X54993; G37066; -.
DR PIR; A34830; A34830.
DR PIR; A34831; A34831.
DR PDB; 1TGH; 01-AUG-96.
DR PDB; 1CDW; 23-DEC-96.
DR MIM; 600075; -.
DR PROSITE; PS00351; TRFID; 2.
DR PFAM; PF00352; TBP; 2.
DR TRANSFAC; T00794; -.
KW TRANSCRIPTION REGULATION; DNA-BINDING; NUCLEAR PROTEIN; REPEAT;
KW 3D-STRUCTURE.
FT DOMAIN 55 95 POLY-GLN.
FT REPEAT 165 241 1.
FT REPEAT 255 332 2.
FT CONFLICT 58 61 MISSING (IN REF. 2).
SQ SEQUENCE 339 AA; 37698 MW; 5D24A7B8 CRC32;

Query Match 2.3%; Score 225; DB 1; Length 339;
Best Local Similarity 38.5%; Pred. No. 1.17e-15;
Matches 42; Conservative 16; Mismatches 49; Indels 2; Gaps 2;

Db 18 QGAMTPGPIPFPMYPGTG-LTPQPIQNTSLSTLEQQRQQRQQRQQRQQRQQRQQRQ 76
Qy 1196 ENPTAGGAVMRPMQPOQGFNAQVQRSR-ELLSHHFQRQVRVMMQMQQQQQQQQQ 1254
Db 77 QQQQQQQQQQQQQQQQQQAAVAAVQOSTQQATQGTSGQAPQLFHSQ 125
Qy 1255 QQQQQQQQQQQQQQQQTQAFSPNPNTASPMQGLLAGTTPQAPPQFPYQ 1303

RESULT 5
ID ARNT DROME STANDARD; PRT; 642 AA.
AC O15945; O16167;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE ARYL HYDROCARBON RECEPTOR NUCLEAR TRANSLOCATOR HOMOLOG (DARNT)
DE (TANGO PROTEIN).
GN TGO OR ARNT.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RC MEDLINE; 98040551.
OHSHIRO T., SAIGO K.;
"Transcriptional regulation of breathless FGF receptor gene by
binding of TRACHEALLESS/darnt heterodimers to three central midline
elements in Drosophila developing trachea.";
RL DEVELOPMENT 124:3973-3986(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RC MEDLINE; 97427859.
RA ZELZER E., WAPPNER P., SHILO B.-Z.;
"the PAS domain confers target gene specificity of Drosophila
bHLH/PAS proteins.";
RT GENES DEV. 11:2079-2089(1997).
RL CC -1- FUNCTION: TGO/TRH HETERODIMERS ARE INVOLVED IN THE CONTROL OF
CC BREATHLESS EXPRESSION.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER
CC BHLH PROTEIN. HETERODIMER WITH AHR, TRH OR SIM.
CC -1- DEVELOPMENTAL STAGE: DISTRIBUTED HOMOGENEOUSLY IN EARLY EMBRYOS,
CC SUGGESTING MATERNAL DEPOSITION OF RNA. AT STAGE 11, EXPRESSION IS
CC DETECTED IN TRACHEAL PITS. AT LATER STAGES, STRONG EXPRESSION IS
CC ALSO DETECTED IN THE CNS.
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB002556; D1023738; -.
EMBL; AF016053; G2367597; -.
DR FLYBASE; FBgn0015014; tgo.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
DR PFAM; PF00010; HLH; 1.
DR PFAM; PF00989; PAS; 1.
KW NUCLEAR PROTEIN; DNA-BINDING; TRANSCRIPTION REGULATION; ACTIVATOR;
KW REPEAT.
FT DNA_BIND 14 26 BASIC DOMAIN.
FT DOMAIN 27 67 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT REPEAT 87 154 PAS-1.
FT REPEAT 273 339 PAS-2.
FT DOMAIN 346 389 PAC MOTIF.
FT DOMAIN 23 26 POLY-ARG.
FT DOMAIN 396 399 POLY-ALA.
FT DOMAIN 404 413 POLY-GLN.
FT DOMAIN 455 462 POLY-GLN.
FT DOMAIN 515 522 POLY-GLN.
FT DOMAIN 566 570 POLY-PRO.
FT CONFLICT 110 110 M -> V (IN REF. 2).
FT CONFLICT 411 411 Q -> QQQ (IN REF. 2).
FT CONFLICT 465 465 R -> G (IN REF. 2).
FT CONFLICT 488 488 P -> T (IN REF. 2).
FT CONFLICT 614 632 EFSDMLQMLDHTPTTFEDL -> GVLRYADVGSADHV
(IN REF. 2).
SQ SEQUENCE 642 AA; 71606 MW; 983D5207 CRC32;

Query Match 2.3%; Score 224; DB 1; Length 642;
Best Local Similarity 34.5%; Pred. No. 1.67e-15;
Matches 50; Conservative 39; Mismatches 42; Indels 14; Gaps 8;

Db 19 CEIERRRNKMTAYITELSDMVPTC-SALAR---KPKLTILRMVAHMKALRGNT-S 73
Qy 31 CSGEKRRREQESKYIELAELISANLSDIDNFNVKPKCALKETVRIQIKQEGKTIS 90
Db 74 SDGIY-KP-SF---LTDQE-LAKHLILEADGFLFVVVSCDSGMIVYSDSVTPVNVQ 125
Qy 91 NDDVQRADVSSTCGQVIDKDSGLGLLQALDGLFVVVNRAN-IVFVSENVQYLTQKQ 149
Db 126 SDWYGTSLYEHHPDREKIREQLS 150
Qy 150 EDLVNTSVYNILHEEDRKDFLKNLP 174

RESULT 6
ID BTD DROME STANDARD; PRT; 644 AA.
AC Q24266;
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE TRANSCRIPTION FACTOR BTD (BUTTONHEAD).
GN BTD.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S; TISSUE=EMBRYO;
RC MEDLINE; 94081952.
RA WIMMER E.A., JAECKLE H., PFEIFLE C., COHEN S.M.;
"A Drosophila homologue of human Spl is a head-specific segmentation
gene.";
RL NATURE 366:690-694(1993).
CC -1- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF THE ANTENNA,
```


[illegible]

QY 1300 FPYQNTYGMQDPAPFAGRVSSPPNAMSRRMGPSQNMQHPQAAISYOSSEMKGWPSG 1359
Db 584 AMNPLNHFHGHHHHH 599
QY 1360 NLARNSFSQOQFAHQ 1375

RESULT 13
ID V192 HUMAN STANDARD; PRT; 2124 AA.
AC Q93074;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN KIAA0192 (FRAGMENT).
GN KIAA0192
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]

SEQUENCE FROM N.A.
TISSUE=BONE MARROW;
MEDLINE: 96281124.
RX NAGASE T., SEKI N., ISHIKAWA K.-I., TANAKA A., NOMURA N.;
RA "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA RES. 3:17-24(1996).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D83783; D1012786; -.
CC KW HYPOTHETICAL PROTEIN.
CC FT NON_TER 1
CC FT DOMAIN 599 602 POLY-SER.
CC FT DOMAIN 1201 1207 POLY-GLY.
CC FT DOMAIN 1998 2124 GLN-RICH.
CC FT DOMAIN 1998 2023 POLY-GLN.
CC FT DOMAIN 2028 2033 POLY-GLN.
CC FT DOMAIN 2037 2070 POLY-GLN.
CC FT DOMAIN 2090 2097 POLY-GLN.
CC SEQUENCE 2124 AA; 237207 MW; 24317D02 CRC32;

Query Match 2.28; Score 213; DB 1; Length 2124;
Best Local Similarity 34.78; Pred. No. 7.89e-14;
Matches 42; Conservative 23; Mismatches 55; Indels 1; Gaps 1;
Db 2003 QQ 2062
QY 1211 PQQGFNAQWVAQRSELLSHFRQORVAMMQQQQQQQQQQQQQQQQQQQQQQQ 1270
Db 2063 HQQQQQQQAAPPQSQFQFQGLQQTQQQQQQQQAALVRQLQQQLSNTQPFSTNIFG 2122
QY 1271 FSPPPNVTASPSMDGLLAGTMTFAPPQPPQPPYQPNYG-MQQPDPAFGRVSSPPNAMS 1329
Db 2123 R 2123
QY 1330 R 1330

RESULT 14
ID HMCU_DROME STANDARD; PRT; 2175 AA.
AC P10150;
DT 01-MAR-1989 (REL. 10, CREATED)
DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)

DE HOMEBOX PROTEIN CUT.
GN CT.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 88232956.
RA BLOCHLINGER K., BODMER R., JACK J., JAN L.Y., JAN Y.N.;
RT "Primary structure and expression of a product from cut, a locus
RT involved in specifying sensory organ identity in Drosophila.";
RL NATURE 333:629-635(1988).
CC -!- FUNCTION: REGULATOR OF CELL FATE DECISIONS IN MULTIPLE LINEAGES.
CC SPECIFICALLY, FUNCTIONS AS A DETERMINATION FACTOR THAT SPECIFIES
CC SENSORY ORGAN IDENTITY IN PRECURSOR CELLS. PROBABLY ALSO INVOLVED
CC IN CELL TYPE SPECIFICATION OF MALPIGHIAN TUBULES. IN ABSENCE OF
CC CUT GENE EXTERNAL SENSORY ORGANS ARE TRANSFORMED INTO CHORDOTONAL
CC ORGANS.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: DETECTED IN MANY CELLS IN THE CENTRAL NERVOUS
CC SYSTEM, ALL EXTERNAL SENSORY ORGANS, SOME PERIPHERAL NEURONS, AND
CC IN THE NON-NEURAL CELLS OF THE SPIRACLES AND THE MALPIGHIAN
CC TUBULES.
CC -!- DEVELOPMENTAL STAGE: CELL-SPECIFIC PATTERN OF EXPRESSION. BROADLY
CC EXPRESSED DURING EMBRYONIC DEVELOPMENT.
CC -!- DOMAIN: ASN AT POSITION 47 OF THE HOMEBOX MAY PARTICIPATE IN
CC REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND
CC HETERODIMERIZATION.
CC -!- SIMILARITY: TO MAMMALIAN COAT DISPLACEMENT PROTEIN (CDP).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X07985; G7768; -.
CC DR EMBL; X07985; E1808; AUT_SEQ.
CC PIR: S03170; S03170.
CC DR FLYBASE; FBgn0004198; ct.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS50071; HOMEBOX_2; 1.
CC DR PFAM; PF00046; homeobox; 1.
CC DR HSP; P04002; IWF.
CC DR TRANSFAC; T02004; -.
CC KW HOMEBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
CC COILED COIL; REPEAT.
CC FT DOMAIN 194 210 ALA/GLN-RICH.
CC FT DOMAIN 235 243 ALA-RICH.
CC FT DOMAIN 271 293 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 384 428 ASN-RICH.
CC FT DOMAIN 433 487 COILED COIL (POTENTIAL).
CC FT DOMAIN 547 554 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 574 584 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 616 630 ALA-RICH.
CC FT DOMAIN 665 699 HIS/GLN-RICH (OPA-REPEAT).
CC FT DOMAIN 886 1689 3 X 'CUT'-REPEATS.
CC FT REPEAT 886 958 'CUT'-REPEAT.
CC FT REPEAT 1339 1411 'CUT'-REPEAT.
CC FT REPEAT 1617 1689 'CUT'-REPEAT.
CC FT DNA_BIND 1745 1804 HOMEBOX.
CC FT DOMAIN 2004 2014 ALA-RICH.
CC FT DOMAIN 2071 2077 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 2124 2136 ALA/PRO-RICH.
CC SEQUENCE 2175 AA; 233628 MW; D7E1FC12 CRC32;

Query Match 2.28; Score 217; DB 1; Length 2175;
Best Local Similarity 23.18; Pred. No. 1.96e-14;
Matches 58; Conservative 77; Mismatches 99; Indels 17; Gaps 14;

This Page Blank (uspto)

WPSRCH_PP

(TM)

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_PP protein - protein database search, using Smith-Waterman algorithm

on: Fri Sep 17 21:12:20 1999; MasPar time 80.13 Seconds
Tabular output not generated. 963.760 Million cell updates/sec

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pep
Perfect Score: 9849
Sequence: 1 MSGLGENLDPLASDRKRKL.....MNNPMPMSGMPGPDQKYC 1415

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 54.229; Variance 118.858; scale 0.456

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	7640	77.6	NUCLEAR RECEPTOR COACT	0.00e+00
2	6892	70.0	RETINOID X RECEPTOR-IN	0.00e+00
3	3346	34.0	TRANSCRIPTIONAL INTERM	0.00e+00
4	3300	33.5	GRIPI.	0.00e+00
5	2887	29.3	NUCLEAR RECEPTOR COACT	0.00e+00
6	1852	18.8	CAGH16.	0.00e+00
7	1856	17.2	STEROID RECEPTOR COACT	2.28e-286
8	1696	17.2	STEROID RECEPTOR COACT	2.28e-286
9	1691	17.2	STEROID RECEPTOR COACT	2.08e-285
10	1674	17.0	NUCLEAR RECEPTOR COACT	3.83e-282
11	1665	16.9	NUCLEAR RECEPTOR COACT	2.05e-280
12	1642	16.7	NUCLEAR RECEPTOR COACT	5.31e-276
13	1643	6.5	STEROID RECEPTOR COACT	1.07e-87
14	376	3.8	(HIN-2).	2.68e-40
15	267	2.7	TPA INDUCIBLE PROTEIN.	3.35e-22
16	232	2.4	AROMATIC HYDROCARBON R	1.06e-16
17	232	2.4	CLOCK (FRAGMENT).	1.06e-16
18	232	2.4	OMEGA SECALIN.	1.06e-16
19	240	2.4	ARYL HYDROCARBON RECEPTOR	6.13e-18
20	240	2.4	ARYL HYDROCARBON RECEPTOR	6.13e-18

21	232	2.4	846	4	015516	CLOCK.	1.06e-16
22	238	2.4	847	6	002747	AH RECEPTOR (ARYL HYDR	1.25e-17
23	235	2.4	855	11	008785	CIRCADIAN LOCOMOTOR OU	3.66e-17
24	233	2.4	1457	5	044011	PROTEIN KINASE YAKA.	7.46e-17
25	223	2.3	304	4	015409	CAGH44 (FRAGMENT).	2.56e-15
26	222	2.3	313	10	041546	ALPHA/BETA-GLIADIN STO	3.63e-15
27	225	2.3	339	4	016845	TRANSCRIPTION INITIATI	1.26e-15
28	230	2.3	357	10	043639	SECI PRECURSOR.	2.16e-16
29	229	2.3	357	10	005573	OMEGA SECALIN PRECURSO	3.08e-16
30	229	2.3	644	5	044082	BASIC HELIX-LOOP-HELIX	6.25e-16
31	229	2.3	720	5	023847	GLUTAMINE-ASPARAGINE R	3.08e-16
32	227	2.3	823	13	073772	AROMATIC HYDROCARBON R	6.25e-16
33	227	2.3	910	11	054899	BRAIN CYCLIC NUCLEOTID	6.25e-16
34	227	2.3	910	11	088704	HYPERPOLARIZATION-ACTI	6.25e-16
35	222	2.3	1307	5	091094	SIMILARITY TO DROSOPHI	3.63e-15
36	226	2.3	2005	4	014686	KIAA0181 PROTEIN (FRAG	8.89e-16
37	228	2.3	2251	5	022190	T05A10.1 PROTEIN.	4.39e-16
38	231	2.3	4957	4	014687	ALR.	1.52e-16
39	231	2.3	5262	4	014686	ALR.	1.52e-16
40	219	2.2	287	10	041509	ALPHA-GLIADIN.	1.04e-14
41	215	2.2	313	10	041529	ALPHA-GLIADIN STORAGE	4.18e-14
42	213	2.2	652	4	015410	CAGH45.	8.35e-14
43	213	2.2	1655	5	024754	MASTERMIND.	8.35e-14
44	213	2.2	2023	4	075557	OPA-CONTAINING PROTEIN	8.35e-14
45	221	2.2	2447	5	022463	SIMILAR TO ZINC FINGER	5.15e-15

ALIGNMENTS

RESULT		1		PRELIMINARY;		PRT; 1398 AA.	
ID	009000	AC	009000;	DT	01-JUL-1997 (TREMBLREL. 04, CREATED)		
DT	01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)			DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	NUCLEAR RECEPTOR COACTIVATOR 3 (P300/CBP/CO-INTEGRATOR PROTEIN).			GN	NCOA3.		
OS	MUS MUSCULUS (MOUSE).			OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;		
OC	SCIUROGNATHI; MURIDAE; MURINAE; MUS.			RN	[1]		
RP	SEQUENCE FROM N.A.			RX	MEDLINE; 97336097.		
RA	TORCHIA J., ROSENFELD M.G.;			RA	TORCHIA J., ROSE D.W., INOSTROZA J., KAMEI Y., WESTIN S., GLASS C.K.,		
RT	"The transcriptional co-activator p/CIP binds CBP and mediates			RT	nuclear-receptor function."		
RL	NATURE 387:677-684 (1997).			RN	[2]		
RP	SEQUENCE FROM N.A.			RA	TORCHIA J., ROSENFELD M.G.;		
RL	SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.			RN	[3]		
RP	SEQUENCE FROM N.A.			RA	TORCHIA J., ROSENFELD M.G.;		
RL	SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			DR	EMBL; AF000581; G2934906; "		
DR	MGD; MGI-1276535; NCOA3.			DR	PFAM; PF00989; PAS; 1.		
DR	PFAM; PF00989; PAS; 1.			SO	SEQUENCE 1398 AA; 151573 MW; 5C9B19B5 CRC32;		
Query Match 77.6%; Score 7640; DB 11; Length 1398;							
Best Local Similarity 79.7%; Pred. No. 0.00e+00;							
Matches 1144; Conservative 146; Mismatches 88; Indels 57; Gaps 14;							
Db	1	MSGLGENSLDPLAASRRKRLPCDAPGGLVYSGEKKRRQESKYIEELAEISANLSDI	60				
Qy	1	MSGLGEN-LDPLASDRKRKLPCDTPGGLTSCGKRRQESKYIEELAEISANLSDI	59				
Db	61	DNFNKPKDKAILKTVRQIQRIKEQKTTSSDDDDVOKADVSSSTGQGVIDKDSGLPLLQ	120				
Qy	60	DNFNKPKDKAILKTVRQIQRIKEQKTTISNDDDDVOKADVSSSTGQGVIDKDSGLPLLQ	119				


```
Db 121 ALDGLFVWNRDGNIVFVSENVTOYLYQKQEDLVNTSVYXILHEPRKDFLNTYQPOLM 180
QY 120 ALDGLFVWNRDGNIVFVSENVTOYLYQKQEDLVNTSVYXILHEPRKDFLNTYQPOLM 179
Db 181 EFLGLMTRDKKAPYLIVRLMKT-HDILEDVNASPEFRQRTETMOCFALSOPRAMLEE 239
QY 180 G-VSWTNEPQORQKSHFNCRMLMKTPHDILEDINASPENRQRYETMOCFALSOPRAMEE 238
Db 240 GEDLOCCMICVARRVTA---PPSPSESFITRDLGSKVNVINDTSLRSMRPFEDDIR 296
QY 239 GEDLOCCMICVARRITGERTFFSPSEFTRDLGSKVNVINDTSLRSMRPFEDDIR 298
Db 297 RCIOREFSLNDGWSQKRRHYQAYVHGHAETPVYRFLADGTIVSAQTKSLFRNPVTN 356
QY 299 RCIOREFSLNDGWSQKRRHYQAYVHGHAETPVYRFLADGTIVTAQTKSLFRNPVTN 358
Db 357 DRHGFISTHFLQREQNGYRPNPTQDKGIRPPAAGC-----GVSMSPNQNVQMGSRITYG 411
QY 359 DRHGFVSTHFLQREQNGYRPNPVGQIRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYG 418
Db 412 VPDPSTNGMGARYGASSVASLTPGOSLOSPSSYONSSYGLSMSSPHGSPGLGNQO 471
QY 419 LADPSTNGMGARYGSSSNIASLTTPGQSPSSYONNIGUNMSSPHGSPGLAPNQ 478
Db 472 NIMISPRNRGSPKMAHQSPVAGVHSPMASSGNTGNHSSSSLSALQAISGVTSL 531
QY 479 NIMISPRNRGSPKIASHQSPVAGVHSPMASSGNTGNHSSSSLSALQAISGVTSL 538
Db 532 STLSGPGKLDNPNPNISOPSKVSQDSKSPGLGLCEQNPVSSVCQNSRDPQVKES 591
QY 539 STLSGPGKLDNPNPNITOPSKVSQDSKSPGLGLCEQNPVSSVCQNSRDLSDKES 598
Db 592 KESSGVESETPRGPLESGHKLLQLLTCSDDRGHSSLTNSPLDNCXDSVSVTSPS 651
QY 599 KESSVEGAENRGPLESGHKLLQLLTCSDDRGHSSLTNSPLDSCRESSVSVTSPS 658
Db 652 VSSSTSGTSSVSNVHGLLQEKHRLHLKLLQNGNSPAEYAKITABETGKDTSTASCE 711
QY 659 VSSSTSGVSSSTNMHGLSLQEKHRLHLKLLQNGNSPAEYAKITAGKDTSTITSCG 718
Db 712 GTT-REQOLSPKKENALLRYLLDRDPDSVLAKELQPADSGDSKLSQCSCSTNPSG 770
QY 719 GNVYQEOQLSPKKENALLRYLLDRDPDSVLAKELQPADSGDSKLSQCSCSTNPSG 778
Db 771 QEKDPKIKTETNDEVGDLNLDAILGLTSSDFYNNP--TNGSHPGAKOOMPAGPSSLG 828
QY 779 QEKDPKIKTETSEGGDLNLDAILGLTSSDFYNNSSSHSLGHTKQVQFGTNSLG 838
Db 829 LRSQPQVQSVRPPYNRAVSLDSPSVSGPPVKNVSAFPGLPKQPIAGNPRMDSQENY 888
QY 839 LKSSQSVQSVIRPPYNRAVSLDSPSVSGPPVKNVSAFPMPLPKQMLGNGNPRMDSQENY 898
Db 889 GANWG-PNRVNPVNTSSPCDGLANSRMRPEPLASSPLGRTGADYSATLPRPAMGGV 947
QY 899 GSSMGPPNRNVYTTQTPSSGDWGLPNSKAGRMEPMNSNGRPGGYNTSLRPAALGSI 958
Db 948 PTLPLRSNRLPGARPSIQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1007
QY 959 PTLPLRSNRLPGARPSIQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO 1000
Db 1008 AVQNSQSPGEGMLMEQCPHGSQNRPLLRNSLDDLPGPPSNAEQSDERALLDLQHTF 1067
QY 1001 AAASNLGSPWDCMLMEQVSHGTQNRPLLRNSLDDLPGPPSNAEQSDERALLDLQHTL 1060
Db 1068 LSNATDGLBEIDRALGIPELVNGQALSKQDFVQGEAAVNMDDKAALYGOTYPAQGP 1127
QY 1061 LSNATDGLBEIDRALGIPELVNGQALEPKQDAFQGEAAVNMDDKAGLYGOTYPAQGP 1120
Db 1128 PLOGGFNLQGSFNSMNGQISQOQSFPLQGMHPRAGLVRPTNTPKOLRWLOLQORLOG 1187
QY 1121 PMOGGFHLQGSFNSMNMNQGNFPLQGMHPRANRPTNTPKOLRWLOLQORLOG 1180
Db 1188 QQFLNRSQALEMKNENP-AG-TAVMRPMN-PO-AFFNAQMAAQKRELMSHHLQOORMA 1243
```

```
QY 1181 QQFLNRSQALEMKNENPTAGGAAYNRPMQPOQGFNAQVQAQRRELLSHFRQORVA 1240
Db 1244 MMMSQPOQP-----AFSPPNVTASPMGDMGVLAGSAMPOAPPQOF 1283
QY 1241 MMQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1300
Db 1284 PYPANYGMOPPEPAFGGSSPPSAMSSRMGSPSONAMVQHQPQTPMYOPSDMKGWPSN 1343
QY 1301 PYQPNYGMQOQPDPAFGRVSSPPNMMSSRMGSPSONPMQHPQAAISYSSSEMKGWPSN 1360
Db 1344 LARNGSFQOQFAPQGNPAAYNNVHMNSGGHGLQWAMTTPMSPGMPGPDQKYC 1398
QY 1361 LARNSFSQOQFAHQGNPAVYSMVHMNGSSGHGMQNMNPMSPGMPGPDQKYC 1415
RESULT 2
ID O57539 PRELIMINARY: PRT: 1391 AA.
AC O57539;
DT 01-JUN-1998 (TREMBREL. 06, CREATED)
DT 01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBREL. 06, LAST ANNOTATION UPDATE)
DE RETINOID X RECEPTOR-INTERACTING COACTIVATOR XICO.
OS XENOPUS LAEVIS (AFRICAN CLAWED FROG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;
OC MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.
RN [1]
RP SEQUENCE FROM N.A.
RA KIM H.-J., LEE S.-K., NA S.-Y., CHOI H.-S., LEE J.W.;
RL MOL. ENDOCRINOL. 0:0-0(1998).
DR EMBL, AF044080; G2852394; -.
SQ SEQUENCE 1391 AA; 152532 MW; C1426F5D CRC32;
Query Match 70.0%; Score 6892; DB 13; Length 1391;
Best Local Similarity 70.0%; Pred. No. 0.00e+00;
Matches 1000; Conservative 239; Mismatches 138; Indels 52; Gaps 27;
Db 1 MSGLENSLDPLASTETKRKRKSSCDTPGGLTCSGKRRRQESKYIEIADLSANLSD 60
QY 1 MSGLEN-LDPLASTSRKKLP-CDTPGGLTCSGKRRRQESKYIEIADLSANLSD 58
Db 61 IDNFNVPDKCAILKETVRIQRIQEQKASSDDVDQKADYSSSTGQGVIDKSLGLPLL 120
QY 59 IDNFNVPDKCAILKETVRIQRIQEQKTSISDDVDQKADYSSSTGQGVIDKSLGLPLL 118
Db 121 QALDGLFYVNRGSGIVFVSENVTOYLYQKQEDLVNTSVYXILHEPRKDFLNTYQPOLM 180
QY 119 QALDGLFYVNRGSGIVFVSENVTOYLYQKQEDLVNTSVYXILHEPRKDFLNTYQPOLM 178
Db 181 NGVPWFSETPROKSHFNCRMLMKTSHDLED-GSNLDARQYETMOCFALSOPRAMEE 239
QY 179 NGVSTNEPQORQKSHFNCRMLMKTSHDLEDINASPENRQRYETMOCFALSOPRAMEE 238
Db 240 GEDLOCCMICVARRITTAERAFSANPESFITRDLGSKVNVINDTSLRSMRPFEDDIR 299
QY 239 GEDLOCCMICVARRITGERTFFSPSEFTRDLGSKVNVINDTSLRSMRPFEDDIR 298
Db 300 RCIOREFLHSEQPTWYKRHYQAYVHGHAETPVYRFLADGTIVTAQTKSLFRNPVTN 359
QY 299 RCIOREFSLNDGWSQKRRHYQAYVHGHAETPVYRFLADGTIVTAQTKSLFRNPVTN 358
Db 360 DRHGFVSTHFLQREQNGYRPNPNAQGITRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYG 418
QY 359 DRHGFVSTHFLQREQNGYRPNPVGQIRPPMAGCNSSVGGMSMSPNOGLQMPSSRAYG 418
Db 417 MGDPSMAQGMRYKSPGNMAPVNAQGVQ-QSPYQNNNSYGLSMSSPHGSPGLGNQO 475
QY 419 LADPSTNGMGARYGSSSNIASLTTPGQSPSSYONN-NGVGLNMSPPHSGPGLAPNQ 477
Db 476 PNLVSPNRASPKWASNOFSPVPGMNSPMGSSGAGGSSSSLSALHAISGEGVSSL 535
QY 478 QNIMISPRNRGSPKIASHQSPVAGVHSPMASSGNTGNHSSSSLSALQAISGEGVSSL 537
```

536 LSSLSPPGKVENNSNMNPOQKICNODCKSPSGLYCEQGVSSVCOSSGREHLGKD 595
538 LSTLSPGPKLDSNMNITQPSKVNQSKSPGLGYCDONPVSSMCSQNSRDLSDKE 597
596 VRENIPEGESORSQAESGKHKLLQLLCTFEERQSLMSSSMD--CKDSS-NVTSPS 652
598 SKRESSVEGAENRGPLESKHKLLQLLCTSSDRGHSSLTNPLDSSCKESSVTSFS 657
653 GYSSSTISGVSTSNLHSGMLQKHLRLKLLONGNSPAEVAKITAEATGKDVQFTVSS 712
658 GYSSSTSGVSTSNMHGSLQKHLRLKLLONGNSPAEVAKITAEATGKDVQFTVSS 715
713 APTCEATV-KREOLSPKKENALLRHLLDKDWPDLAKIDKPKVEHMDIKMGSCSSN 771
716 --CGDNVYKQELSPKKENALLRYLLDRDPDALSKELOPQVEGVNDKMSQCTST 773
772 VPTSSODKEVKTETEPGVPGLDNLDAIPLGAGSDFYNSMSRASDLGPKQPVFOD 831
774 IPSSEKEDPKIKETSEEGSGDLNDAIPLGAGSDFYNSMSRASDLGPKQPVFOD 833
832 SPTLAMRSPDSMGSRRPPFNRAMSLDSR-S--STPPRVNVSFPLPKQMGIGS-PRMMD 887
834 TNSLGLKSSQSVQIRPPYNAVSLDPSVSGSSPPVKNISAFPLPKQPMGLGNPRMD 893
888 GQDNFVGMGSGPNR--SMNQHPG-GDMQNSAVNRLEPPNVGSGVRGPGDYSSAWTRP 944
894 SQENYSSMG-GPNRVNVTQTPSSGDWGLPNSKAGRMEMPNSMGRPGDYNTSLPRP 952
945 AMGNMPLGLTNSNTPGSPVWQOQHILPMRPNDMAMSGSNPYGOQAPSNPPGSPWD 1004
953 ALGSGTPTPLRSNIPGAPRVQOQOQMLQMRPGEIPMGGANPYGOAAASNLQSGWPD 1012
1005 AIM-MNQGRCAGNOLGRNSLDLCLPSTVGEQTDALQLLHLLTNTDGLGEEI 1063
1013 GMLSMQVSHGTQNRPLRLNSLDLVGPPSNLESGSDERALLQLLHLLTNTDGLGEEI 1072
1064 DRALGTPDLVQSQALPEQDPSDTPQSGSPVWIDQKPMYQGHYAGGAAAGSAGFNMQG 1123
1073 DRALGIPELVNOGQALEPKODAFQGGQAAVMDQKAGLYQGYTPAOGPPM-QGGEH-LQ 1130
1124 QHPENTVMQMOQOQGMHPLOQGMHPRLNIRPNIPKOLRMOLQORLOGQOFLQNRQ 1183
1131 QSPFSNMMNQO-QGNFLOGMHPRLNIRPNIPKOLRMOLQORLOGQOFLQNRQ 1189
1184 ALBWKYDPMNPGGAVRPMVQTPVSOQFLNAQVMAQKRELISHQIRQHRMAMMQOQ 1243
1190 ALEKMEPTAGGAAYRPMQ-P-QQFLNAQVMAQKRELISHQIRQHRMAMMQOQ 1246
1244 QGQFQ-----AFSPPNVTASMDNPLGPPMPQAPQPPQFPPNY 1285
1247 QQQQQQQQQQQQQQQQQQQQQQAFSPPNVTASMDNPLGPPMPQAPQPPQFPPNY 1306
1286 GINQOTDPTGVRSSPPNAMSRRMAPSONP---HPQOTOMYSPDMKMGPSGNMARPN 1342
1307 GMGQDPDPAFGVRSSPPNAMSRRMAPSONP---HPQOTOMYSPDMKMGPSGNMARPN 1366
1343 FPOQVYSHOTNPATYNMHNMGNGHGMQVNSLPSMGPMGPDQKYC 1391
1367 FSOQFAHQONPAYSVNHMGSSGHHGMQVNSLPSMGPMGPDQKYC 1415

RESULT 3
ID Q15596 PRELIMINARY; PRT; 1464 AA.
AC Q15596;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE TRANSCRIPTIONAL INTERMEDIARY FACTOR 2.
GN TIF2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE-PLACENTA;
RX MEDLINE; 96312964.
RA VOGEL J.J., HEINE M.J.S., ZECHEL C., CHAMON P., GRONMEYER H.;
RT "TIF2, a 160 kDa transcriptional mediator for the ligand-dependent
activation function AF-2 of nuclear receptors.";
RL EMBO J. 15;3667-3675(1996).
DR EMBL; X97674; E307031; -.
DR PFAM; PF00989; PAS; 1.
SQ SEQUENCE 1464 AA; 159156 MW; C761519D CRC32;

Query Match 34.0%; Score 3346; DB 4; Length 1464;
Best Local Similarity 44.1%; Pred. No. 0.00e+00;
Matches 653; Conservative 370; Mismatches 318; Indels 139; Gaps 84;

Db 1 MSGMGENTSDPSRAETRRK-ECPDQGPSKRNTEKRNREKENYIEELAEILAFNFD 59
QY 1 MSGLGENL-DPLASDSRKRKLP-CTPGGLTCSGKRRREKESYIEELAEILANLSD 58
Db 60 IDNFNPKDKAILKETVQIRKEQEKAAANIDEVOKSDVSSTGQGVIDKDALGPM 119
QY 59 IDNFNPKDKAILKETVQIRKEQEKISND-DVQKADVSSTGQGVIDKDALG 117
Db 120 LEALDGFVFNLEGNVTVSENVTOYLYNOEELMKNKSVSYLLHVGHDHTEFVK 179
QY 118 LQALDGFVFNLEGNVTVSENVTOYLYNOEELMKNKSVSYLLHVGHDHTEFVK 176
Db 180 IYNGSGSEPPRRNSHTFCNMLVPLPDSEEGHNDNOEAHOKYETMOCFAVSQPK 239
QY 177 TVNGSVTNEPQKSHTEFCNMLVPLPDSEEGHNDNOEAHOKYETMOCFAVSQPK 236
Db 240 EGEDLQSLICVARVPMKRPVLPSSFTTRDQLQKITSLDSTWRAAMKPGWEDL 299
QY 237 EGEDLQSLICVARVPMKRPVLPSSFTTRDQLQKITSLDSTWRAAMKPGWEDL 296
Db 300 VRCIOKFAHQEGESVYAKHHHEVLROGLAFSIOYRESLSDGTLVAAOTKSLIR 359
QY 297 IIRCIQRTFLSDQGSW-QKRHYQAYLNHAEETPVYFLADGTIVTAQTKSLIR 355
Db 360 TTNEPQLVSLHMLHREQVNVNPNDLTGTMGKPLNIPSSNSPAHQALCSGNPG 419
QY 356 VINDRGVSTHEFLQRCQCYRPNPVGQI-RP-PMAGCNSSVGM-SMSPNOGLQ 411
Db 420 SSNINFPINGPKQMGPMGRFGSGGMNHVS---GQATTP-QGSYALKMNSPSS 475
QY 412 PSSRAYGLADPSTTGQMSGARVGGSNIASTLPGCMQSPSYQNNYGLNMSPPH 471
Db 476 GMPGQPTSLSPRHRMSPGACVPRIQSPQSPAGSLHSPVGVCSSTGNSHST 535
QY 472 GLAPNQNTMISPRN-----GSPKIASHQFSPVAGVHSPMASSNGTGN-HS 524
Db 536 ALQALSEGCHGVSLGSLASPLDKMNLQNSPVNMPPLSLKMGSLDSCDGLY-GE 593
QY 525 ALQALSEGCHGVSLGSLASPLDKMNLQNSPVNMPPLSLKMGSLDSCDGLY-GE 590
Db 594 EGTGQAESCHPGEQKTNPNLPAVSSERADGOSRLHDSKQTKLLQLLTSSQ 653
QY 581 ESSMCSQNSRDLSD-KESKESVEGA-ENQ-R-G-P-L-ESKGHKLLQLLTSS 633
Db 654 PSPLASS-LSDTNKDSGLS--P-G-----SG---ST---HGTSLKHKILH 698
QY 634 HSLTNSPLDSSCKESVSVTSFSGVSSSTSGVSSSTNNMHGSLQKHLRLK 693
Db 699 SPVDLAKLTAEATGKDLQSSSTAPGSEVTIKOEPVSPKKEN-ALLRYLLDK 757
QY 694 SPAEVAKITAEATGKDLQSSSTAPGSEVTIKOEPVSPKKEN-ALLRYLLDK 749
Db 758 GLP-EITPKLERLDSKTDPAASNTKLIAMTEKE-EMSFEPGQPGSELDN 815
QY 750 ALSKELQPVQGVGVDNKMSSQCTSTIPSSSQEKDPKIKETSEEGSGDLN 808
Db 816 SOLPQLFPDTPRGAAGVSDKQAIINDLMOLTAENSPVTPVGAQKTAIRLS 875
QY 816 SOLPQLFPDTPRGAAGVSDKQAIINDLMOLTAENSPVTPVGAQKTAIRLS 875

QY 809 S--SDFYNNSIS-S-NGS---H--LGT-KQ---Q---VFQ-GTNSLGLKSSQSV-QSIRP 850
Db 876 GQLGRLLPNLPLDITLQSPGTGAGPEPIRNSPVSVIPQPMGMG-NOGIMIGNOGLGN 934
QY 851 -----PY-N-R-AVLDSPVSGSSPPVRNIFAPMLPKQPMGLGGRNPMDSQENTGS 900
Db 935 SGTGIMGNASRPTMPSGEWA-POSSAVRVTCATTSAMNRPV-Q-G-GMIRNP-AA5IP 989
QY 901 SMGGPNRVTVTQTPSSGDWGLPNSKACRME-PMNSNSMGRPGDYNTSLRPPALGG5IP 959
Db 990 ---MRPSSQPGQRQLQSV-N-NIGPSELENNMGPGPOYSQQOAPPNTATWPESILPID 1044
QY 960 TLPLRSNIPGARPYLOOQOQMLQRPGEIPMGMGANPYGA-AA5NOLG5WPDGML5ME 1018
Db 1045 QASFA5QNRQPPGSPDCLCPHPAAE5P5DEGALLDOLYLALRFD--GLEEIDRALGI 1102
QY 1019 QVSHGTQNRPLRLN5LDDLVGPP5NEQ5DERALLDQHL5L5NTD5T5G5EIDRALGI 1078
Db 1103 PELVSQ5QAVDPEQ--F5SQDSNIMLEQAPVFPQOYASQAQMAQGSYS-PMQDPNFHTM 1159
QY 1079 PELVNOGQALEPKDADFQGEAAVMDQKAGLYGQTYPAQGP5MG5FHLOG5SP5NSM 1138
Db 1160 -CQ--RPSYATILR-MQPRPG-LRPTGLVQNPQNLRLQLQHLQAOQ--N--RQPL-MNQ 1209
QY 1139 MNQMNQOQGNFPLQGMHPRANIMRPR--T-NTPKQLRMQLQORLQOQQLN5RQALELKM 1195
Db 1210 ISNV5NVLTLRPGV-PTQAPINAO5LQOR5EILNHLRQ-R-QM-----HQ 1254
QY 1196 ENPTAGGAAV5M5MPQOQGE5LNAQ5V5AQR5RELL5H5FRQ5V5AM5MQOQOQOQOQ 1255
Db 1255 QOQVQOQRLMRGQGLN5T5M5V5P5G5M5P5M5N5P5R5I5Q5A5Q5P5P5P5N5Y5G5I5Q5O5P5D5G 1314
QY 1256 QOQ 1315
Db 1315 FTGATTPQ5L5P5M5A5H5T5Q5M5Q5O5Q5A5N5P5A5Q5D5I5N5G5A5Q5N5F5S5Q5S5P5H 1374
QY 1316 FQV5S5P5P5N5M5S5R5M5G5P5O5N5P5M5H5Q5A5I5Q5-S5M5K5W5P5N5L5A5R5N5S5F5S5Q5--Q 1371
Db 1375 FQOQANT5M5N5M5N5M5N5M5T5G5M5S5M5N5Q5M5T--GQISM 1413
QY 1372 FAHQ5N5P5A5V5M5-VH5N5G5-SGH5M5G5M5N5M5P5M5G5M5P5M 1408

RESULT 4
ID Q61026; PRELIMINARY; PRY; 1462 AA.
AC Q61026; P97759;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
GRIPI.
GN GRIPI.
OS MUS MUSCULUS (MOUSE).
OC EURKYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
CC SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RC STRAIN=ICR OUTBRED; TISSUE=BRAIN;
RP SEQUENCE OF 322-1119 FROM N.A.
RX MEDLINE; 96209838.
RA HONG H., KOHLI K., TRIVEDI A., JOHNSON D.L., STALLCUP M.R.;
RT "GRIPI, a novel mouse protein that serves as a transcriptional
coactivator in yeast for the hormone binding domains of steroid
receptors.";
RT PROC. NATL. ACAD. SCI. U.S.A. 93:4948-4952(1996).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR OUTBRED; TISSUE=BRAIN;
RX MEDLINE; 97265407.
RA HONG H., KOHLI K., GARABEDIAN M.J., STALLCUP M.R.;
RT domain of steroid, thyroid, retinoid, and vitamin D receptors."
RL MOL. CELL. BIOL. 17:2735-2744(1997).
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=ICR OUTBRED; TISSUE=BRAIN;
RA STALLCUP M.R.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR OUTBRED; TISSUE=BRAIN;
RA HONG H., STALLCUP M.R.;
RL SUBMITTED (FEB-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EXBL; U39060; G1853980; --
DR MGD; MGI:107429; GRIPI.
DR PFAM; PF00989; PAS; I.
SQ SEQUENCE 1462 AA; 158511 MW; 9F9EC6A7 CRC32;

Query Match 33.5%; Score 3300; DB 11; Length 1462;
Best Local Similarity 42.8%; Pred. No. 0.00e+00;
Matches 631; Conservative 375; Mismatches 339; Indels 129; Gaps 81;

Db 1 MSGGENTSDPSRAETRRK--ECPDQLGSPKPRSTERRNRQENKYIEELADLIFANFD 59
QY 1 MSGLGENL-DPLASDSRRKLPK-DTPCQGLTCSGKRRRQESKYIEELAEELISANLSD 58
Db 60 IDNENFKPKCAILKETVKQIRKKEKAAANIDEVQKSDSVSTGGVTDKDALGPM 119
QY 59 IDNENFKPKCAILKETVKQIRKKEKQKTSND--DDYQKADSVSTGGVTDKDSLGPLL 117
Db 120 LEALDGGFFVYNLEGSVYFVSENVTOYLRYNOEELMKSVY5ILHVGDHTEFVNKLPS 179
QY 118 LQALDGGFLVYNREANIVFVSENVTOYLQKQEDLVNTSVNIIHEEDRKDFLKNL-PKS 176
Db 180 MVNGSGSGEPRTSTFTNCRLMLVKPLPDSEEGHGSQEAHOKYEQAMQFAV5QPKSIX 239
QY 177 TVNGSVTNEPQKSHFTNCRLMLKPHDLIEDINASPEMRQRYETMQCFAL5QPRAMM 236
Db 240 EGEDLQSLCIVARRVPMKERPTLPSS-ESFTTRDQLOGKITSLDSTMRAMKPGWED 298
QY 237 EGEDLQSLCIVARRITGER-TFSPNPE5FITRHLDSGKVVNIDTNSLRSSMRPGFD 295
Db 299 LVRCIQKFTQHEGESLSYAKRHHHEVLROGLAF5IYREFSLSDGTLVAQTKSLIRS 358
QY 296 IIRCIQRF5SLNDGQ5WS-QKRHYEAYLNGHAETPVYREF5ADGTVTVAOTKSLFRN 354
Db 359 QTTNEPOLV5LHMLHREQNV5VMNPDLTGQAMGKPLNP5SSSP5HAQ5L5G5NPGQDMT 418
QY 355 PVTNDRHGF5V5THFLQREQNGYRPNPNVQGI-RP--PMAGCN5V5GGM-SM5PNOGLQ 410
Db 419 LGSNINFP5MNGPK5QM5MP5G5G5G5M5N5V5--QKQATTP-QGSN5ALKM5SP5OSS 474
QY 411 MP5SRAYGLADP5T5TG5M5G5A5R5G5S5N5I5A5LT5P5G5M5Q5SP5Y5QNNY5GLN5SP5PHGS 470
Db 475 PGMNPGQASSV5L5PRQ5M5PGV5AG5PRIP5Q5F5P5AG5L5HP5G5V5CS5T5G5N5SH5YT5N5SL 534
QY 471 PGLAPNQNIM5ISPNR-----G5PKIASHQF5PV5AG5V5H5P5MA5SS5GNT5GN-H5F5SS5L 523
Db 535 NALQAL5EGH5V5L5SG5S5L5AS5DL5K5M5GL5Q5N5P5M5N5P5P5L5K5M5GL5DK5C5F5LY-G5-P 592
QY 524 SALQAISEG5V5T5L5L5S5SP5GK--LDNSP-NMNITQ5PK5V5N5Q5D5K5PL5G5YC5QNP 579
Db 593 SEGTTQ5A5E5A5CH5PE5OK5P5ND5SM--PQA5GD-RAG5H5RL-H--DSK5QTK5LL5QLLT 646
QY 580 VESSM5Q5NS5RD5H5LD--KESK5E5VEGA5N5Q5R5L5K5H5K5K5L5QLLT5C5SD5DR5H5SLT 638
Db 647 TKS-DQM-EP5PLP-SSI5D5TK5D5T5G5L5P5G5T5H5C5T5L5K5E5K5H5L5R5L5D5SS5VDL 703
QY 539 NSPLD5CK5E5SV5T5P5G5V5SS5T5G5V5ST5N5H5G5L5Q5E5K5H5L5K5L5Q5N5S5PA5EY 698
Db 704 AKLTAEATG5L5Q5E5SS5T5AP5G5E5TV5K5Q5E5P5AS5PK5KEN-ALL5R5LL5DK5D5TK5IG5P-E 761
QY 699 AKITAQATG5D5T5S5IT5SGD-GN-V-VK5Q5L5SP5KK5EN5ALL5R5L5D5R5D5PSD-ALSKE 754
Db 762 ITP5KL5R5L5D5K5T5D5PAS5TK5IAM5TV5KE-EV5F5P5Q5P5Q5E5L5N5E5L5D5L5Q5N5LPQ 820
QY 755 LQ5QV5G5D5N5K5M5S5Q5T5S5T5P5SS5Q5E5K5D5PK5IK5T5E5G5D5L5N5D5AIL5D5L-TS--SD 811

Db 821 LFPDTRGAPTGSVDKQAIINDLMOLTADSPVPVPPAGAAKAALRMSTFNNRPRQGLGR 880
 QY :
 812 FYNNIS-S-NGS--H-LGT-KQ--Q--VFQ-GTNSLGLKSSQSV-QSIRP----- 850

Db 881 LLPNQLPLDIITLQPTGAGPEPPILRNSSPSVLPQPGMG-NQMLGSGOGLGNNTGM 939
 QY :
 851 --PY-N-R-AVSLDSPVSVGSSPPVKNISAFPMPLPKPOMLGGPRMWDSONEYSSMGPG 905
 Db 940 IGSTSRPSWPGGEWA-POSFAVRVTCAATGAMNRPV-Q-G-GMRNP-TASIP---MR 991
 QY :
 906 NRNVTVTOTPSSDGWLGNPKSKRGME-PMNSNSMGRGGDYNTSLRPALGSGIPTLR 964
 Db 992 ANSQCGORMLASOV-M-NIGPSELENNMGPOYNQQAAPNPOTAPMPESILPIDQASPA 1049
 QY :
 965 SNSIPGARVLVQQOOQLMRPGEIPEMGGANPYGQA-AAASNOLGSPDCMLSMSEQVSHG 1023

Db 1050 SONRQPFSGSPDDLLCPHPAESPSDEGALLDOLYLALRNF-D-GLEETDRALGIPELYS 1107
 QY :
 1024 TQNRPLRLNSLDLVGPSPNLGEGSDERALLDQHLLTSNTDATGLEIDRALGIPELVN 1081

Db 1108 QSQAIVDAEQ-FSSQESSIMLEQKPVPVPOQYASQAQMAOGGYN-PMQDPNFHTM-GQ-- 1161
 QY :
 1084 QGALEPKODAQGOEAAMVMDOAKLAGLYGVYPAGGPPMGGHFHLOQSFSNMNQMN 1143

Db 1162 RPNYTTLR-MQPRPGLRPTGIIVQNOPNLRLOLQHRLOAQO--N-RQPL-MNQISSVSN 1215
 QY :
 1144 QQGNFFLOGMHPRANIMRP-RT-NTPKQLRMQLQRLQGQOFLNQRSAOLELKMENPTAG 1201

Db 1216 VNLTLRPGV-PTQAPINAQMLAQRRETLNHQRO-R-QM-----QQQVOO 1258
 QY :
 1202 GAAMVPMRPMPOQGFNLNAQMAQRSRELLSHHFQQRVAMMOQQQQQQQQQQQQQQ 1261

Db 1259 RTLMRGGLINTVTPSNAPAGLPAAMSNPRIPOANAQOFPPPNYGISQOPDPGFTGATT 1318
 QY :
 1262 QQOQOQTQAFSPPNVTASPSMDGLAGTWPQPAPPQFPQPNYMGQOPDPAFORVSS 1321

Db 1319 PQSPLMSPRMAHTQSPMMQOSQANPAYOPTSDMGNWAGSGMGNMFSSQSPPHFGQCAN 1378
 QY :
 1322 PPNAMSSRRMGPSONPMOHPOQAASIYQ-SSEMKGWPSGNLARNSSFSQO---QFAHQGN 1377

Db 1379 TSMTYNNMNIYSMATNGLSLNNMQT-COMS 1411
 QY :
 1378 PAYISM-VHMINGS-SCHMGQM-NMPPMSGNPM 1408

RESULT 5 PRELIMINARY; PRT; 1463 AA.
 O09001;
 O09001;
 01-JUL-1997 (TREMBREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBREL. 04, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
 DE NUCLEAR RECEPTOR COACTIVATOR 2
 DE (NUCLEAR RECEPTOR COACTIVATOR PROTEIN 2).
 GN NC0A2.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 SC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97336097.
 RA TORCHIA J., ROSE D.W., INOSTROZA J., KAMEI Y., WESTIN S., GLASS C.K.,
 RT ROSENFIELD M.G.;
 RT "The transcriptional co-activator p/CIP binds CBP and mediates
 nuclear-receptor function";
 RL NATURE 387:677-684(1997).
 DR ENBL; AF000582; G2213817; -
 DR MGD; MG1:1276533; NC0A2.
 DR PFAM; PF00989; PAS; 1.
 SQ SEQUENCE 1463 AA; 158880 MW; E367B699 CRC32;

Query Match 29.3%; Score 2887; DB 11; Length 1463;
 Best local similarity 40.1%; Pred. No. 0.00e+00;
 Matches 590; Conservative 380; Mismatches 377; Indels 26; Gaps 81

```

1052 QNRQPTGSSPDDLCPHPAAESPDGALLDQYLALRNF--GLEIDRALGIPELVSQ 1109
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1025 QNRLPRLNSLDLVGSPNLEGOSDERALLDLQHTLLSNTDATGLEIDRALGIPELVNQ 1084
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1110 SQAVDAQ--FSSQESSIMLEQPPVFPQOYASQAQAAGGYN--PMQDPNPHM--GQ--R 1163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1085 GQALEPKQDAFOGQEAANVMQDKAGLYGQTPAQGPPMOGGFHHGQSPFSFNMMNMNQ 1144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1164 PNVTTLR-MQPRPGLRPTGIVQNPNQLRLQLQHLRAQ--N--ROPL-MNOISSVSVN 1217
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1145 QGNFPLQGHPRANIRP--RT--NTPQLRMQLQRLQGLQGLNOSRALEKLKXENPTAGG 1202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1218 NLTLRGV--PTQAPINQAQLQRREILNOHLRQ--R-QM-----QQQVQQR 1260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1203 AAVMRMMPMQQGLFNAQVAQSRRELLSHHFROQRVAMMMQOQOQOQOQOQOQOQ 1262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1261 TLMRGOGLNVTPSVYAPAGLPAAMSNPRIQANAQOFPFPNTYGISQDPDPGFTGNTP 1320
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1263 QOQQOQTAFSPPNVTASPSMDGLAGTMTQPAPQFPYQPNYGMQOQPDPAFGRVSSP 1322
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1321 QSLMSPRAHQTQSPMNOQANLAYQPTSDMNGWAQSGMGNMFWSQSPPHFHGQOANT 1380
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1323 PNAMSSRWGSPONPMQHPQASIIY--SSEMKGHPSONLARNSSFSQO---QFAHQGNP 1378
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1381 SMYSN--NMNISVSMATNTGGLSSMNQMTGQMSM 1412
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1379 AVYSVMYHMGSSGHMGOMN--MNP--PMG--NPM 1408
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6
ID 015406 PRELIMINARY; PRT; 326 AA

AC	Q15406;
AD	01-JAN-1998 (TREMBREL. 05, CREATED)
AE	01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
AF	01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
AG	CAGH16.
AH	CAGH16.
AI	HOMO SAPIENS (HUMAN).
AJ	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
AK	CATARRHINI; HOMINIDAE; HOMO.
AL	[1]
AM	SEQUENCE FROM N.A.
AN	TISSUE=BRAIN;
AO	MEDLINE; 97369492.
AP	RA MARGOLIS R.L., ABRAHAM M.R., GATCHELL S.B., LI S.H., KIDWAI A.S.,
AQ	RA BRESCHER T.S., STINE O.C., CALLAHAN C., MCINNIS M.G., ROSS C.A.;
AR	"CDNAS with long CAG trinucleotide repeats from human brain.";
AS	HUM. GENET. 100:114-122(1997).
AT	EMBL; U80737; G2565050; -.
AV	SEQUENCE 326 AA; 36669 MW; FE1786F6 CRC32;

Query Match 18.8%; Score 1852; DB 4; Length 326;
Best Local Similarity 84.1%; Pred. No. 0.00e+00;
Matches 275; Conservative 12; Mismatches 25; Indels 15; Gaps 7;

Db	1	MMDQKAGLGYQTYPAQGPPMQGFGFLHQGSFSFNSMNMNQGNFFPLQGHPRANTMRP	60
Qy	1103	MMDQKAGLGYQTYPAQGPPMQGFGFLHQGSFSFNSMNMNQGNFFPLQGHPRANTMRP	1162
Db	61	RTNTPQLRMQLQRLRQLGQQOFLNQSRQALELKLMENPTAGGAAYVRPMQMOPVSSQEGEFLN	120
Qy	1163	RTNTPQLRMQLQRLRQLGQQOFLNQSRQALELKLMENPTAGGAAYVRPMQMOPQ--Q-GFLN	1218
Db	121	AQMGQRRELLSHHFRRQVRAMMQGQQGQQGQQGQQGQQGQQGQQGQQGQQEQEVES	180
Qy	1219	AQMVQRSRELLSHHFRRQVRAMMQGQQGQQGQQGQQGQQGQQGQQGQQGQQ-----TQAFS	1272
Db	181	PPPNVTAPAWMAXGRTHNATSSSATSTQYP-TPNYNGDQPTKIQLPGRVSKSSQCNDV	239
Qy	1273	PPPNVTASPS-MD-GLLAGTTPMPAQPOOFPYQPNYGMGOFPD--PAGFRVSFPFNMMMS	1328
Db	240	VKNPGSQNPMMQHQPAASTIQSSEMKGWPSGNLARNSFSQOQFAHOGCNPAVTSWYHNG	299

[illegible]

Query Match 17.2%; Score 1696; DB 4; Length 1398;
Best Local Similarity 40.0%; Pred. No. 2.28e-286;
Matches 339; Conservative 223; Mismatches 217; Indels 68

Db	1	SGLGSSDDPANPDSPHRRKSPCDT-----LASTERRRRQENKYEEELAEELSANISDI	56
Qy	2	SGLGENT-DPLASRRRK-LPDCDTPGGITCGEKKRRREQESKYIEELAEELSANISDI	59
Db	57	DSLGVKPKCKILKKTVDQQLMKRMQEKSTT-DDDVQKSDTSSSSQGVIEKSLGPLL	115
Qy	60	DNFNVKPKCALLKTTVYRQIK--EQGKTSINDDDVQKADVSSGQGVIDRDSLGPLL	117
Db	116	LEALDGEFFVYVNCBGRIVFVSENVTSLYGVNQBELMNTSVSYILHYGDHAEFYKNLLPKS	175
Qy	118	LQALDGLFLVYNREANVFVSENVTQYLYQKQEDLVNTSVYTLHEEDRKDFKLN--PKS	176
Db	176	LVNGVVPQEQATRENSTHFNCRMLIHPD-D--EPGTENQACQRYEVMQCFTVSYQPSKIQ	232
Qy	177	TVNGSVNTNPEQKSHTFNCRMLKTPHDILEDINASPEMRQRYETMQCFALQSPAMM	236
Db	233	EDGEDFQSLCITARRLPSP-PAI-TGVESFMTKQDTTKIISDTSLSRAAGRTGWEDL	290
Qy	237	EEGEDLQSCMICVARRITTCGERTFPSPESFIRPHDLGKGVINDTNSLRSLRMRPGFEDI	296

Db	349	
Db	349	
Qy	356	
Db	409	
Qy	408	
Db	469	
Qy	462	
Db	529	
Qy	515	
Db	598	
Qy	571	
Db	641	


```

Db 57 IDSLSVKPKCKILKTKVTDIOQLMKRMEQEKSTT-DDDVQKSDISSSQGVIEKESLGLP 115
QY IDNFNVPKPKCKILKTKVTDIOQLMKRMEQEKSTT-DDDVQKSDISSSQGVIEKESLGLP 116
Db 116 LLEALDGFVFNCEGRIVFVSENVSYLYGNOBELMNTSVSYLHVGDHAEFVKNLLPK 175
QY LLEALDGFVFNCEGRIVFVSENVSYLYGNOBELMNTSVSYLHVGDHAEFVKNLLPK 175
Db 176 SLVNGVPMQEAQRRNSHTFNCMLKTHPP-D--EPGTENQEAQRYVQMCFTVSQPKSI 232
QY SLVNGVPMQEAQRRNSHTFNCMLKTHPP-D--EPGTENQEAQRYVQMCFTVSQPKSI 232
Db 176 STVNGVSWTNEPQRQKSHFTFNCMLKTHPP-D--EPGTENQEAQRYVQMCFTVSQPKSI 232
QY STVNGVSWTNEPQRQKSHFTFNCMLKTHPP-D--EPGTENQEAQRYVQMCFTVSQPKSI 232
Db 233 QEDGEDFQSCILCIARRLRP-PAT-TGVESEMTKQDTTKIISIDTSLRAAGRTGWD 290
QY QEDGEDFQSCILCIARRLRP-PAT-TGVESEMTKQDTTKIISIDTSLRAAGRTGWD 290
Db 236 MEEGEDLQSCMICVARRITGERTFSPNPESTFTRHDLGSKVNVNIDTSLRSMRPGFD 295
QY MEEGEDLQSCMICVARRITGERTFSPNPESTFTRHDLGSKVNVNIDTSLRSMRPGFD 295
Db 291 LVKRCIYAFQPGQREPSVAROLFQEVMTGRTASSPSVREILNDGTMLSAHTKCKL-CYP 349
QY LVKRCIYAFQPGQREPSVAROLFQEVMTGRTASSPSVREILNDGTMLSAHTKCKL-CYP 349
Db 296 IIRRCIQRFSLNDGQSWQKRYQAYLNGHAETPVYRFLADGTIVTAQTKSLFRNP 355
QY IIRRCIQRFSLNDGQSWQKRYQAYLNGHAETPVYRFLADGTIVTAQTKSLFRNP 355
Db 350 QSPDMQPFIMGIIHIDREHSGLSPODDTNGSMISIPRVNPSVNPISPAHGVARSSTLPSP 409
QY QSPDMQPFIMGIIHIDREHSGLSPODDTNGSMISIPRVNPSVNPISPAHGVARSSTLPSP 409
Db 356 VTNDRGFV-STHFLQREQNGYRPNP-VGQI-R--PPMAGCNSV-G-GMSMS-PNQ 407
QY VTNDRGFV-STHFLQREQNGYRPNP-VGQI-R--PPMAGCNSV-G-GMSMS-PNQ 407
Db 410 NSNVSTRINROQSSDLHSSSHSNSSQSGFCSPGQITVANVALNKGQASSQSKPSL 469
QY NSNVSTRINROQSSDLHSSSHSNSSQSGFCSPGQITVANVALNKGQASSQSKPSL 469
Db 408 GLQMPSSRAY-GLA-D-PPSTQMGARYGGS-S-NIAS-LTPGPGMQ-SPSSVQNNYGL 461
QY GLQMPSSRAY-GLA-D-PPSTQMGARYGGS-S-NIAS-LTPGPGMQ-SPSSVQNNYGL 461
Db 470 NLNPMWEGTGISLAQFMSPPROVTSGLATPRMPNPPNPSTLSSPVGMTSSACNN 529
QY NLNPMWEGTGISLAQFMSPPROVTSGLATPRMPNPPNPSTLSSPVGMTSSACNN 529
Db 462 NMSGPP-HGSP-GLAP-NOQINMISPRNRSKPIASHQFSP-VAGVHSP--MASSG-NTG 514
QY NMSGPP-HGSP-GLAP-NOQINMISPRNRSKPIASHQFSP-VAGVHSP--MASSG-NTG 514
Db 530 NRSYSNIPVTSLOGMEGNPNVSGFASPSVLRQMSQSPSLNI-QPAKAESKDKEI 588
QY NRSYSNIPVTSLOGMEGNPNVSGFASPSVLRQMSQSPSLNI-QPAKAESKDKEI 588
Db 515 NRSYSNIPVTSLOGMEGNPNVSGFASPSVLRQMSQSPSLNI-QPAKAESKDKEI 588
QY NRSYSNIPVTSLOGMEGNPNVSGFASPSVLRQMSQSPSLNI-QPAKAESKDKEI 588
Db 589 ASPL-NEM-IQSDNSSDQK-PL-DSGLLHNDRLSDGO-SKYSQTSK-LVOLIT-TTA 641
QY ASPL-NEM-IQSDNSSDQK-PL-DSGLLHNDRLSDGO-SKYSQTSK-LVOLIT-TTA 641
Db 571 LGFYCDQNPVSSCMQSNRDLSDKESKESVGEAGNQRGLSKHKLQLLTCSDD 630
QY LGFYCDQNPVSSCMQSNRDLSDKESKESVGEAGNQRGLSKHKLQLLTCSDD 630
Db 642 EQ-Q--LRHADITSDKDV-LSCGTGTSNANSSGSGCPSS-HSSL-TARHKLRLHQL 695
QY EQ-Q--LRHADITSDKDV-LSCGTGTSNANSSGSGCPSS-HSSL-TARHKLRLHQL 695
Db 631 DRGHSSLTNSPLDSSCKESSVTSVSPGVSSTSGVSSSTNNMHGSLQKHLHKLQ 690
QY DRGHSSLTNSPLDSSCKESSVTSVSPGVSSTSGVSSSTNNMHGSLQKHLHKLQ 690
Db 696 EG-SPDITLSPEDPKOSASTSVTVQVQGNSSIKLELDASKKESKHOLLYLID 754
QY EG-SPDITLSPEDPKOSASTSVTVQVQGNSSIKLELDASKKESKHOLLYLID 754
Db 691 NGSNPAEVAKITAAQTKDTSIT-SC-GD--GNV-VKQEQJSPKKE--NNALLRYLID 743
QY NGSNPAEVAKITAAQTKDTSIT-SC-GD--GNV-VKQEQJSPKKE--NNALLRYLID 743
Db 755 KDEKDLRSTPNLSDDDVKKVKEKQMDPCNTNPTMTKTPP-EIKLEPAQSOFTADLQ 813
QY KDEKDLRSTPNLSDDDVKKVKEKQMDPCNTNPTMTKTPP-EIKLEPAQSOFTADLQ 813
Db 744 RDPSPDALSKELO-PQVEG-VDNK--MSQCTSTIPSSQEKDPKIKTETSEBEGSDLN 799
QY RDPSPDALSKELO-PQVEG-VDNK--MSQCTSTIPSSQEKDPKIKTETSEBEGSDLN 799
Db 814 FDQLLPTL 821
QY FDQLLPTL 821
Db 800 LDAILGDL 807
QY LDAILGDL 807

```

RESULT 10

```

ID P70366 PRELIMINARY; PRT: 1405 AA.
AC P70366;
DT 01-FEB-1997 (TREMBREL. 02, CREATED)
DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE NUCLEAR RECEPTOR COACTIVATOR 1 (STEROID RECEPTOR COACTIVATOR-1).
GN NCOAL OR MSRC-1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97008053.
RA YAO T.P., KU G., ZHOU N., SCULLY R., LIVINGSTON D.M.;

```

```

RT "The nuclear hormone receptor coactivator SRC-1 is a specific target
of p300."
RL PROC. NATL. ACAD. SCI. U.S.A. 93:10626-10631(1996).
DR EMBL; U64828; G1490876; .
DR MGD; MGI:1276523; NCOAL.
DR PFAM; PF00989; PAS; 1.
SQ SEQUENCE 1405 AA; 152643 MW; 1743F755 CRC32;

```

Query Match 17.0%; Score 1674; DB 11; Length 1405;

```

Best Local Similarity 39.8%; Pred. No. 3 83e-282;
Matches 339; Conservative 231; Mismatches 212; Indels 70; Gaps 55;

```

```

Db 1 MSGLDSSDNPANPSHKKSGSPCDT----LASSTKRRREQENKYLELAELLANISD 56
QY IDNFNVPKPKCKILKTKVTDIOQLMKRMEQEKSTT-DDDVQKSDISSSQGVIEKESLGLP 116
Db 57 IDSLSVKPKCKILKTKVTDIOQLMKRMEQEKSTT-DDDVQKSDISSSQGVIEKESLGLP 115
QY IDNFNVPKPKCKILKTKVTDIOQLMKRMEQEKSTT-DDDVQKSDISSSQGVIEKESLGLP 116
Db 116 LLEALDGFVFNCEGRIVFVSENVSYLYGNOBELMNTSVSYLHVGDHAEFVKNLLPK 175
QY LLEALDGFVFNCEGRIVFVSENVSYLYGNOBELMNTSVSYLHVGDHAEFVKNLLPK 175
Db 176 SLVNGVPMQEAQRRNSHTFNCMLKTHPP-D--EPGTENQEAQRYVQMCFTVSQPKSI 232
QY SLVNGVPMQEAQRRNSHTFNCMLKTHPP-D--EPGTENQEAQRYVQMCFTVSQPKSI 232
Db 233 QEDGEDFQSCILCIARRLRP-PAT-TGVESEMTKQDTTKIISIDTSLRAAGRTGWD 290
QY QEDGEDFQSCILCIARRLRP-PAT-TGVESEMTKQDTTKIISIDTSLRAAGRTGWD 290
Db 236 MEEGEDLQSCMICVARRITGERTFSPNPESTFTRHDLGSKVNVNIDTSLRSMRPGFD 295
QY MEEGEDLQSCMICVARRITGERTFSPNPESTFTRHDLGSKVNVNIDTSLRSMRPGFD 295
Db 291 LVKRCIYAFQPGQREPSVAROLFQEVMTGRTASSPSVREILNDGTMLSAHTKCKL-CYP 349
QY LVKRCIYAFQPGQREPSVAROLFQEVMTGRTASSPSVREILNDGTMLSAHTKCKL-CYP 349
Db 296 IIRRCIQRFSLNDGQSWQKRYQAYLNGHAETPVYRFLADGTIVTAQTKSLFRNP 355
QY IIRRCIQRFSLNDGQSWQKRYQAYLNGHAETPVYRFLADGTIVTAQTKSLFRNP 355
Db 350 QSPDMQPFIMGIIHIDREHSGLSPODDTNGSMISIPRVNPSVNPISPAHGVARSSTLPSP 409
QY QSPDMQPFIMGIIHIDREHSGLSPODDTNGSMISIPRVNPSVNPISPAHGVARSSTLPSP 409
Db 356 VTNDRGFV-STHFLQREQNGYRPNP-VGQI-R--PPMAGCNSV-G-GMSMS-PNQ 407
QY VTNDRGFV-STHFLQREQNGYRPNP-VGQI-R--PPMAGCNSV-G-GMSMS-PNQ 407
Db 410 NLNPMWEGTGISLAQFMSPPROVTSGLATPRMPNPPNPSTLSSPVGMTSSACNN 529
QY NLNPMWEGTGISLAQFMSPPROVTSGLATPRMPNPPNPSTLSSPVGMTSSACNN 529
Db 408 GLQMPSSRAY-GLA-D-PPSTQMGARY-G--GSSNIASLTTPGPGM-QSPSSVQNNYGL 460
QY GLQMPSSRAY-GLA-D-PPSTQMGARY-G--GSSNIASLTTPGPGM-QSPSSVQNNYGL 460
Db 470 NLNPMWEGTGISLAQFMSPPROVTSGLATPRMPNPPNPSTLSSPVGMTSSACNN 529
QY NLNPMWEGTGISLAQFMSPPROVTSGLATPRMPNPPNPSTLSSPVGMTSSACNN 529
Db 461 NMSGPP-HGSP-GLAP-NOQINMISPRNRSKPIASHQFSP-VAGVHSP--MASSG-NTG 514
QY NMSGPP-HGSP-GLAP-NOQINMISPRNRSKPIASHQFSP-VAGVHSP--MASSG-NTG 514
Db 530 NRSYSNIPVTSLOGMEGNPNVSGFASPSVLRQMSQSPSLNI-QPAKAESKDKEI 588
QY NRSYSNIPVTSLOGMEGNPNVSGFASPSVLRQMSQSPSLNI-QPAKAESKDKEI 588
Db 514 GNHSFSSSLALQISEGVGTSLLTSSPGKLDNSPNMNTQPSKYSNODSKSPG 573
QY GNHSFSSSLALQISEGVGTSLLTSSPGKLDNSPNMNTQPSKYSNODSKSPG 573
Db 586 SKEIASILNEMTQSDNSD-SANESKPLDGLHNDRLSEGDGSKYSQTSKHLVQLLT-T 643
QY SKEIASILNEMTQSDNSD-SANESKPLDGLHNDRLSEGDGSKYSQTSKHLVQLLT-T 643
Db 574 YCDQNPVSSCMQSNRDLSDKESKE-SSVEGAENQR-GPLESK-GHK--KLQLLTCS 628
QY YCDQNPVSSCMQSNRDLSDKESKE-SSVEGAENQR-GPLESK-GHK--KLQLLTCS 628
Db 644 TAEQ-Q--LRHADITSDKDV-LSCGTGTSNANSSGSGCPSS-HSSL-TARHKLRLH 697
QY TAEQ-Q--LRHADITSDKDV-LSCGTGTSNANSSGSGCPSS-HSSL-TARHKLRLH 697
Db 629 SDRGHSSLTNSPLDSSCKESSVTSVSPGVSSTSGVSSSTNNMHGSLQKHLHKL 688
QY SDRGHSSLTNSPLDSSCKESSVTSVSPGVSSTSGVSSSTNNMHGSLQKHLHKL 688
Db 698 LOEG-SPDITLSPEDPKOSASTSVTVQVQGNSSIKLELDASKKESKHOLLYLID 756
QY LOEG-SPDITLSPEDPKOSASTSVTVQVQGNSSIKLELDASKKESKHOLLYLID 756
Db 689 LQNGNSPAEVAKITAAQTKDTSIT-SC-GD--GNV-VKQEQJSPKKE--NNALLR 739
QY LQNGNSPAEVAKITAAQTKDTSIT-SC-GD--GNV-VKQEQJSPKKE--NNALLR 739
Db 757 YLLDKEDLRLSTPNLSDDDVKKVKEKQMDPCNTNPTMTKTPP-EIKLEPAQSOFTA 815
QY YLLDKEDLRLSTPNLSDDDVKKVKEKQMDPCNTNPTMTKTPP-EIKLEPAQSOFTA 815
Db 740 YLLDRDPSDALSKELO-PQVEG-VDNK--MSQCTSTIPSSQEKDPKIKTETSEBEG 795
QY YLLDRDPSDALSKELO-PQVEG-VDNK--MSQCTSTIPSSQEKDPKIKTETSEBEG 795
Db 816 DLDDQFQLLPTL 827
QY DLDDQFQLLPTL 827
Db 796 DLDDQFQLLPTL 827
QY DLDDQFQLLPTL 827

```



```

RESULT 11
ID Q61202; PRELIMINARY; PRT; 1405 AA.
AC Q61202;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE NUCLEAR RECEPTOR COACTIVATOR 1 (STEROID RECEPTOR COACTIVATOR 1A).
GN NCOAL OR SRCIA.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA KAMEI Y., XU L., HEINZEL T., TORCHIA J., KUROKAWA R., GLOSS B.,
RL LIN S.C., HEYMAN R.A., ROSE D.W., GLASS C.K., ROSENFELD M.G.;
RL CELL 0:0-0:0.
DR EMBL; U56920; G1336160; -.
DR MGD; MGI:1276523; NCOA1.
PFAM; PF00989; PAS; 1.
SEQUENCE 1405 AA; 152571 MW; 670CDD71 CRC32;

Query Match 16.9%; Score 1665; DB 11; Length 1405;
Best Local Similarity 39.7%; Pred. No. 2,05e-280;
Matches 338; Conservative 229; Mismatches 215; Indels 70; Gaps 55;

Db 1 MSGIGSSSDPANPDHSHKRGKSPCDT----LASTKRRRQENKYLEGLAELLSANISD 56
QY 1 MSGIGENL-DPLASDSKRK-LPCDTPGQGLTCGKRRRQESKYIEELAEELISANLSD 58
Db 57 IDSLSVKPKCKILKKTVDQIQLKMRKEQKSTT-DDDVQKSDISSSQGVIEKESLGPL 115
QY 59 IDNFNVPKPKCALIKETVROIRK--EOGKTTISNDVQKADVSSGTQGVIDKDSLGLPL 116
Db 116 LLEALDGFVFNCEGRIVFVSENVTSYLYGNOELMNTSVSYLHVGDHAEVKNLLPK 175
QY 117 LLEALDGFVFNCEGRIVFVSENVTSYLYGNOELMNTSVSYLHVGDHAEVKNLLPK 175
Db 176 SLVNGVWPQAEATRRNSHTFNCRLIHPPED--PGTE-NOEACQRYEVMQFTVSPQKSI 232
QY 176 STVNGSVTNEPQOKSHTEFNCRLMKTPHDILEDINASPEMRQRYETMQCFALSQPRM 235
Db 233 QEGEDFQSLCIAARLPRP-PAI-TGVESFMTKQDTTKIISIDTSSRAAGRTGWED 290
QY 236 MERGEDLQSCMICVARRITGERTFPSPESFTRHDLGKVNIDTNSLRSMRPGFED 295
Db 291 LVKRCIYAFPOQGRPSYAROLFQEVMTGTASSPSYRFLNDGTMLSAHTCKKL-CYP 349
QY 296 IIRRCIORFSLNDGQSWOKRHYQEAYLNGHAETPVYRFLADGTIVTAQTKSKLFRNP 355
Db 350 QSPDMQPFIMGIIHIDREHSGLSQDDNSGMSIPRINPSVNPFGISPAGVTRSTLPPS 409
QY 356 VTNDRHGFV-STHFLQREQNGYRPNP-VQGGI-R--PPM-AGCNSSVG-GMSMS-PNQ 407
Db 410 NNNMVSARVNRQSSDLNSSSHTSNNQNGFCGPGNOIVANVALNQOAGSQTNPS 469
QY 408 GLQMPSSRAY-GLA-D-PSTTGOMSGARY-G--GSSNIASTLTPGGM-QSPSSYQNNNYG 460
Db 470 LNLNNSPMEGTGALSQFMSPRQANSGLATRAMSNNSFPNPTLSSPVGITSACNN 529
QY 461 LNMSSPPHSGPLAPNQNTIMISPRNG--S-PKIASHQFSP-VAGVHSPMA--SSG--NT 513
Db 530 NNRYSNIPVTSLQGMNEGNNSV-G-FSAGSPVLRQMSQNS-S-PSRLSMQPAKESKD 585
QY 514 GNHSFSSSLALQAISEGVTSLTSSPGPKLDNSPNNTITQPKSVNQDSKSLGF 573
Db 586 SKETASILNMIQSDNSD-SANEGKPLDGLLHNDRLSEGDSKYQTSKHLVQLLT-T 643
QY 574 YCDQNPVEMSSCMQNSRDHLSDKESKE-SSVEGAENQR-GPLESK-GHK-KLIQLLTC 628
Db 644 TAEQ-Q--LRHADITCKDV-LSCTGTSSASNSPSSGTCPPSS-HSSL-TERKHLIHLR 697
QY 629 SDDRGHSSLTNSPLDSSCKESSVTSVSPGVSSTSGVSTSNMHGSLLOEKHRIHLK 688

```

```

Db 698 IQEG-SPSDITLTSVEPEKDSVPASTAYSVSGOSGASIKLELDAKKESKDQHLR 756
QY 689 LONGNSPAEVAKITAQATGKDT--SSIT-S-CGD--GNV-VKQEQSLPKKKE--NNALLR 739
Db 757 YLLDKDEKDLRSTPNCLDDVKVKEKQMDPCNTNPTMTKPAPE-EVKLESQSOFTA 815
QY 740 YLLDRDDPSDALSKEL-QPOVEG-VDNK--MSOCTSTTSPSSQEKDPKIKETSEGG 795
Db 816 DLDQFDQLLPTL 827
QY 796 DLDNLDAILLGDL 807

RESULT 12
ID P70365; PRELIMINARY; PRT; 1447 AA.
AC P70365;
DT 01-FEB-1997 (TREMREL. 02, CREATED)
DT 01-FEB-1997 (TREMREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE NUCLEAR RECEPTOR CO-ACTIVATOR.
GN MNRG-1.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA ZHU Y., QI C., CALANDRA C., RAO M.S., REDDY J.K.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U64606; G1490872; -.
DR PFAM; PF00989; PAS; 1.
SEQUENCE 1447 AA; 156896 MW; A92E6387 CRC32;

Query Match 16.7%; Score 1642; DB 11; Length 1447;
Best Local Similarity 39.4%; Pred. No. 5.31e-276;
Matches 336; Conservative 231; Mismatches 215; Indels 70; Gaps 55;

Db 1 MSGIGSSSDPANPDHSHKRGKSPCDT----LASTKRRRQENKYLEELAEELLSANISD 56
QY 1 MSGIGENL-DPLASDSKRK-LPCDTPGQGLTCGKRRRQESKYIEELAEELISANLSD 58
Db 57 IDSLSVKPKCKILKKTVDQIQLKMRKEQKSTT-DDDVQKSDISSSQGVIEKESLGPL 115
QY 59 IDNFNVPKPKCALIKETVROIRK--EOGKTTISNDVQKADVSSGTQGVIDKDSLGLPL 116
Db 116 LLEALDGFVFNCEGRIVFVSENVTSYLYGNOELMNTSVSYLHVGDHAEVKNLLPK 175
QY 117 LLEALDGFVFNCEGRIVFVSENVTSYLYGNOELMNTSVSYLHVGDHAEVKNLLPK 175
Db 176 SLVNGVWPQAEATRRNSHTFNCRLIHPPED--PGTE-NOEACQRYEVMQFTVSPQKSI 232
QY 176 STVNGSVTNEPQOKSHTEFNCRLMKTPHDILEDINASPEMRQRYETMQCFALSQPRM 235
Db 233 QEGEDFQSLCIAARLPRP-PAI-TGVESFMTKQDTTKIISIDTSSRAAGRTGWED 290
QY 236 MERGEDLQSCMICVARRITGERTFPSPESFTRHDLGKVNIDTNSLRSMRPGFED 295
Db 291 LVKRCIYAFPOQGRPSYAROLFQEVMTGTASSPSYRFLNDGTMLSAHTCKKL-CYP 349
QY 296 IIRRCIORFSLNDGQSWOKRHYQEAYLNGHAETPVYRFLADGTIVTAQTKSKLFRNP 355
Db 350 QSPDMQPFIMGIIHIDREHSGLSQDDNSGMSIPRINPSVNPFGISPAGVTRSTLPPS 409
QY 356 VTNDRHGFV-STHFLQREQNGYRPNP-VQGGI-R--PPM-AGCNSSVG-GMSMS-PNQ 407
Db 410 NNNMVSARVNRQSSDLNSSSHTSNNQNGFCGPGNOIVANVALNQOAGSQSNPS 469
QY 408 GLQMPSSRAY-GLA-D-PSTTGOMSGARY-G--GSSNIASTLTPGGM-QSPSSYQNNNYG 460
Db 470 LNLNNSPMEGTGALSQFMSPRQANSGLATRAMSNNSFPNPTLSSPVGITSACNN 529
QY 461 LNMSSPPHSGPLAPNQNTIMISPRNG--S-PKIASHQFSP-VAGVHSPMA--SSG--NT 513
Db 530 NNRYSNIPVTSLQGMNEGNNSV-G-FSAGSPVLRQMSQNS-S-PSRLSMQPAKESKD 585

```


Oy 1288 AGPTMPQAPPQQFPYQPNYGMGQQDPDPAFGRVSSPPNMMSSRMGP-SQN-PMQHPOA 1344

Search completed: Fri Sep 17 21:16:25 1999
Job time : 245 secs.

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: September 17, 1999, 20:42:28 ; Search time 58.07 Seconds
(without alignments)
7097.947 Million cell updates/sec

Title: US-09-041-994-1
Perfect score: 4496
Sequence: 1 GCTGGATGGTGGACTCAGAG.....CATTGACGAGGAATTCTAG 4496
Scoring table: IDENTITY_NUC
Searched: 176461 seqs, 45838279 residues
Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	73.2	1.6	195	3	US-08-469-802B-2
2	73.2	1.6	234	3	US-08-469-802B-3
3	73.2	1.6	168	3	US-08-469-802B-4
4	73.2	1.6	171	3	US-08-469-802B-5
5	73.2	1.6	195	4	US-08-267-803B-2
6	73.2	1.6	224	4	US-08-267-803B-3
7	73.2	1.6	168	4	US-08-267-803B-4
8	73.2	1.6	171	4	US-08-267-803B-5
9	68.4	1.5	3366	3	US-08-469-802B-6
10	68.4	1.5	506	3	US-08-469-802B-7
11	68.4	1.5	506	3	US-08-469-802B-8
12	68.4	1.5	3366	4	US-08-267-803B-1
13	68.4	1.5	134	4	US-08-267-803B-6
14	68.4	1.5	506	4	US-08-267-803B-7
15	68.4	1.5	10660	4	US-08-267-803B-8
16	67.6	1.5	10366	2	US-08-246-982A-5
17	67.6	1.5	10366	2	US-08-453-265-5
18	67.6	1.5	10366	4	US-08-457-273B-41
19	66.4	1.5	3376	1	US-08-320-559-29
20	66.4	1.5	3376	5	PCT-US94-04496-29
21	64	1.4	379	3	US-08-145-617-5
22	63.2	1.4	1776	4	US-08-531-927B-1
23	59.6	1.3	325	4	US-08-531-927B-3
24	59.4	1.3	4835	3	US-08-386-495-9
25	59.4	1.3	4835	5	PCT-US96-02331-9
26	56.6	1.3	1332	4	US-08-481-814A-3
27	56.4	1.3	5063	3	US-08-185-432-1
28	56.4	1.3	3771	3	US-08-185-432-3
29	56.4	1.3	3771	3	US-08-185-432-23
30	56.2	1.2	3819	1	US-07-686-322A-3
31	56.2	1.2	3819	3	US-08-002-999-3
32	56.2	1.2	7498	4	US-08-816-693A-1
33	55.2	1.2	3231	2	US-08-195-152-1
34	54	1.2	32207	4	US-08-770-379-20
35	52.8	1.2	2946	2	US-08-473-122-1
36	52.8	1.2	1518	4	US-08-820-170A-20
37	52.8	1.2	2636	4	US-08-820-170A-21

38 52.8 1.2 2946 4 US-08-472-478-1 Sequence 1, Appli
39 52.8 1.2 2946 4 US-08-463-081B-7 Sequence 7, Appli
40 52.8 1.2 2289 4 US-08-463-081B-30 Sequence 30, Appli
41 52.8 1.2 2946 4 US-08-461-379A-7 Sequence 7, Appli
42 52.8 1.2 2289 4 US-08-461-379A-30 Sequence 30, Appli
43 52.8 1.2 2946 4 US-08-462-390B-7 Sequence 7, Appli
44 52.8 1.2 2289 4 US-08-462-390B-30 Sequence 30, Appli
45 52.6 1.2 2950 5 PCT-US93-08386-7 Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-469-802B-2
; Sequence 2, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muehling, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muehling, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-2

Query Match 1.6%; Score 73.2; DB 3; Length 195;
Best Local Similarity 69.7%; Pred. No. 1.7e-11;
Matches 99; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 3768 AGCTGCTAGTCACTCTCCGACACAGAGGGTGTATGATGTCACGACGACGAC 3827
Db 44 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 103
QY 3828 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3887
Db 104 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 163
QY 3888 CCCAGGCCCTTACGCCACCTCC 3909
||| | ||||| |||||

	Query Match	1.68;	Score 73.2;	DB 3;	Length 171;
	Best Local Similarity	69.74;	Pred. No. 1.6e-11;		
Matches	99; Conservative	0;	Mismatches 43;	Indels	0; Gaps 0;
QY	3768	AGCTGCTAAGTCATCACTTCCGCACAACAGAGGGTTGGCTATGATGATCAGCAGCAGCACAC	3827		
Dd	20	AGC	79		
QY	3828	AGCAGCAGCAGCAGCAGCAGCAGCAGCAAACAGCAACAGCAACAGCAGCAACAGCAGCAAA	3887		
Dd	80	AGC	139		
QY	3888	CCNAGGCCCTTCAGCCCCAGCTCCC	3909		
Dd	140	AGCAGCACCTCAGCAGGGGTCCC	161		

```

RESULT 6
US-08-267-803B-3
Sequence 3, Application US/08267803B
Patent No. 5834183
GENERAL INFORMATION:
APPLICANT: Ori, Harry T.
APPLICANT: Rannum, Laura P.W.
APPLICANT: Chung, Ming-yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
Patent No. 5834183
TITLE OF INVENTION: Type 1 and Method for Diagnosis
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Muetling, Raasch, Gebhardt & Schwappach, P.A.
STREET: P.O. Box 581415
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55458-1415
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/267,803B
FILING DATE: 28-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McCoirmack, Myra H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00030120
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO. 3:

```


Db 1481 CAAATGGGCGAGTCTGAGCCAGACGCCGGGACACAAAGGCTGAGCAGCAGCAGCAGCA 1540
QY 3820 GCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 3879
Db 1541 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1600
QY 3880 GCAGCAAAACCCAGGCCCTTCAGCCCACTCC 3909
Db 1601 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1630

Search completed: September 18, 1999, 01:49:11
Job time: 18403 sec

Result No.	Query	Score	Query			ID	Description
			Match	Length	DB		
1	4451.6	99.0	6835	1	V99915	AIb1 (Amplified in	
2	2812.2	62.5	4821	1	V99919	Murine pCIP gene (
3	2750.8	61.2	4860	1	X26000	Murine pCIP gene.	
4	561	12.5	6156	1	V03517	Human transcriptio	
5	432.4	9.6	462	1	V44857	Clone AJ1.1 5' cod	
6	392.4	8.7	418	1	V90274	EST clone D1454_N	
7	261.8	5.8	325	1	V85262	EST clone AJ1. New	
8	76.2	1.7	1310	1	V08558	Transgene for epit	
9	76.2	1.7	4286	1	V08559	Transgene for epit	
10	76.2	1.7	3263	1	V08560	Transgene for epit	
11	73.2	1.6	195	1	O84831	Spinocerebellar at	
12	73.2	1.6	234	1	O84832	Spinocerebellar at	
13	73.2	1.6	168	1	O84833	Spinocerebellar at	
14	73.2	1.6	171	1	O84834	Spinocerebellar at	
15	68.6	1.5	645	1	T42476	HuntL4 coding sequ	
16	68.6	1.5	2294	1	V80181	Alternatively splic	
17	68.6	1.5	3334	1	V80117	Osf2/cbfa1 polypep	
18	68.4	1.5	3366	1	O84803	Spinocerebellar at	
19	68.4	1.5	306	1	O84804	Spinocerebellar at	
20	68.4	1.5	10660	1	O84793	Spinocerebellar at	
21	68.4	1.5	154	1	O84835	Spinocerebellar at	
22	67.8	1.5	203	1	V17226	SCA2 gene CAG repe	
23	67.8	1.5	203	1	V30271	Glutamine rich reg	
24	67.6	1.5	10366	1	O67401	Huntingtin DNA\Ttl	
25	67.6	1.5	10366	1	T97524	Composite sequence	
26	67.6	1.5	10366	1	V05828	Human huntingtin c	
27	67.6	1.5	515	1	X18930	Human huntingtin g	
28	67.4	1.5	165	1	V30274	Glutamine rich reg	
29	66.8	1.5	4367	1	V30270	Gene causative of	
30	66.6	1.5	1065	1	O74149	Plasmid psi4001 al	
31	66.6	1.5	1680	1	O74150	Plasmid psi4001 al	
32	66.4	1.5	3376	1	T975166	AF-9 cDNA. New acu	
33	65.4	1.5	4200	1	T78912	Spinocerebellar at	
34	65.4	1.5	516	1	V06551	SCA2 gene fragment	
35	65.4	1.5	4481	1	V06552	Human SCA2 cDNA in	
36	64	1.4	379	1	O04602	Cryptically simple	
37	62.8	1.4	3569	1	N91772	Human androgen rec	
38	62.8	1.4	3569	1	T63407	Androgen receptor	
39	62.6	1.4	2069	1	X22940	Mouse calpain CAPN	
40	62.4	1.4	3217	1	N91578	Rat androgen recep	
41	62.4	1.4	4180	1	N91773	Rat androgen recep	
42	62.4	1.4	3217	1	Q12002	Full-length rat an	
43	62.2	1.4	78	1	V12732	SCA2 gene CAG repe	

Db 356 ATCTGCCAATCTTAGTGATATTGACAAATTTCAATGTCAAAACCAGATAAAATGTGCGATTTT 415
Qy 301 AAAGGAACAGTAGACAGATACGTCAAAATAAAGACAGAGGARAACACTATTTCCCAATGA 360
Db 416 AAAGGAACAGTAGACAGATACGTCAAAATAAAGACAGAGGARAACACTATTTCCCAATGA 475
Qy 361 TGATGATGTTCAAAAAGCCGATGTATCTCTACAGGCGAGGAGTTATTGATAAAGACTC 420
Db 476 TGATGATGTTCAAAAAGCCGATGTATCTCTACAGGCGAGGAGTTATTGATAAAGACTC 535
Qy 421 CTTAGGACCGGTTTACTTCAGGCATTGGAGTGTTCCTATTGTTGGTGAATCGAGAGCC 480
Db 536 CTTAGGACCGGTTTACTTCAGGCATTGGAGTGTTCCTATTGTTGGTGAATCGAGAGCG 595
Qy 481 AATCATGTTATGTAACAGAAATGTCACACAATACCTGCAATATAGCAAGAGGACCT 540
Db 596 AATCATGTTATGTAACAGAAATGTCACACAATACCTGCAATATAGCAAGAGGACCT 655
Qy 541 GGTTAACACAAAGTGTTCACAAATATCTTACATGAAGAAGACAGAAAGGATTTTCTTAAGAA 600
Db 656 GGTTAACACAAAGTGTTCACAAATATCTTACATGAAGAAGACAGAAAGGATTTTCTTAAGAA 715
Qy 601 TTTACCAAAATCTACAGTTTAATGGAGTTTCTGGACAAATGAGCCCCCAAGACAAAAG 660
Db 716 TTTACCAAAATCTACAGTTTAATGGAGTTTCTGGACAAATGAGCCCCCAAGACAAAAG 775
Qy 661 CCATACATTAATTTGCCGTATGTTGATGAACACACACATCATATCTGCAAGACATAA 720
Db 776 CCATACATTAATTTGCCGTATGTTGATGAACACACACATCATATCTGCAAGACATAA 835
Qy 721 CGCCAGTCTGGAATGGCCAGAGATATGAACAAATGCAGTGTTCGCCCTGTCTAGCC 780
Db 836 CGCCAGTCTGGAATGGCCAGAGATATGAACAAATGCAGTGTTCGCCCTGTCTAGCC 895
Qy 781 ACAGCTATGATGAGGAAGGGAAGATTGCAATCTTGTATGATCTGTGTGGCACGCCG 840
Db 896 ACAGCTATGATGAGGAAGGGAAGATTGCAATCTTGTATGATCTGTGTGGCACGCCG 955
Qy 841 CATTACTACAGGAAGAAACATTTCCATCAAAACCTGAGAGCTTTATTACCGACATGA 900
Db 956 CATTACTACAGGAAGAAACATTTCCATCAAAACCTGAGAGCTTTATTACCGACATGA 1015
Qy 901 TCTTTCAGGAAGGTTGTCAATATAGATACAAATTCACAGATCCCTCCATGAGGCCGTG 960
Db 1016 TCTTTCAGGAAGGTTGTCAATATAGATACAAATTCACAGATCCCTCCATGAGGCCGTG 1075
Qy 961 CTTTGAAGATATAATCCGAAGGTGTTACAGAGATTTTGTAGTCTAAATGATGGCAGTC 1020
Db 1076 CTTTGAAGATATAATCCGAAGGTGTTACAGAGATTTTGTAGTCTAAATGATGGCAGTC 1135
Qy 1021 ATGCTCCAGAAACGTCACATCAAGAAGCTTATCTTAATGGCCATCGAANAACCCAGT 1080
Db 1136 ATGCTCCAGAAACGTCACATCAAGAAGCTTATCTTAATGGCCATCGAANAACCCAGT 1195
Qy 1081 ATATCGATTTCTGTTGGCTGATGGAATATAGTGACTGACAGACAAAAGCAAACTCTT 1140
Db 1196 ATATCGATTTCTGTTGGCTGATGGAATATAGTGACTGACAGACAAAAGCAAACTCTT 1255
Qy 1141 CCGAAATCTGTATAAATAATGATCGACATGGCTTTGTCTCAACCCACATCTCTCAGAGAGA 1200
Db 1256 CCGAAATCTGTATAAATAATGATCGACATGGCTTTGTCTCAACCCACATCTCTCAGAGAGA 1315
Qy 1201 ACAGAAATGGATATAGACCAACCAATCTGTTGGACAGGATAGACCACTATGGC 1260
Db 1316 ACAGAAATGGATATAGACCAACCAATCTGTTGGACAGGATAGACCACTATGGC 1375
Qy 1261 TGGATGCAACAGTTTCGGTAGCGGCATGAGTATGTCGCCAAACCAAGGCTTACAGATGCC 1320
Db 1376 TGGATGCAACAGTTTCGGTAGCGGCATGAGTATGTCGCCAAACCAAGGCTTACAGATGCC 1435
Qy 1321 GAGCAGAGGCGCTTATGGCTTGGCAGACCCCTAGCACACAGGCGAGATGAGTGAGCTAG 1380
Db 1436 GAGCAGAGGCGCTTATGGCTTGGCAGACCCCTAGCACACAGGCGAGATGAGTGAGCTAG 1495

Qy 1381 GTATGGGGTTCAGTAACATAGCTTTCATTGACCCCTGGGCCAGGCATGCAATCACCATC 1440
Db 1496 GTATGGGGTTCAGTAACATAGCTTTCATTGACCCCTGGGCCAGGCATGCAATCACCATC 1555
Qy 1441 TTCTTACCAGAACCAACACTATGGGCTCAACATGAGTAGCCCCCACATGGGAGTCTG 1500
Db 1556 TTCTTACCAGAACCAACACTATGGGCTCAACATGAGTAGCCCCCACATGGGAGTCTG 1615
Qy 1501 TCTTGCCCAAAACACAGAGATATCATGATTTCTCTCTGTAATCGTGGGAGTCCAAAGAT 1560
Db 1616 TCTTGCCCAAAACACAGAGATATCATGATTTCTCTCTGTAATCGTGGGAGTCCAAAGAT 1675
Qy 1561 AGCCTCACATCAGTTTCTCTCTGAGTGTGACACTCTCCCATGSCATCTTCTGSCAA 1620
Db 1676 AGCCTCACATCAGTTTCTCTCTGAGTGTGACACTCTCCCATGSCATCTTCTGSCAA 1735
Qy 1621 TACTGGAAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCATAGTGAAG 1680
Db 1736 TACTGGAAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCATAGTGAAG 1795
Qy 1681 TGTGGGACTTCCCTTTATCTACTCTGTCATCACCAGGCCCAAAATTTGGATTAACCTCTCC 1740
Db 1796 TGTGGGACTTCCCTTTATCTACTCTGTCATCACCAGGCCCAAAATTTGGATTAACCTCTCC 1855
Qy 1741 CAATATGAATATTACCAACCAAGTAAAGTAAGCAATCAGGATTTCCAAAGAGTCTCTG 1800
Db 1856 CAATATGAATATTACCAACCAAGTAAAGTAAGCAATCAGGATTTCCAAAGAGTCTCTG 1915
Qy 1801 CTTTATTGGACCAAAATCCAGTGGAGAGTTCAATTTGTCTCAGTCAAAATAGCAGAGATCA 1860
Db 1916 CTTTATTGGACCAAAATCCAGTGGAGAGTTCAATTTGTCTCAGTCAAAATAGCAGAGATCA 1975
Qy 1861 CTTCACTGACAAAGTAAGTAAAGAGAGCAGTGTGTGAGGGGCGAGAGAAATCAAGAGGCTCC 1920
Db 1976 CTTCACTGACAAAGTAAGTAAAGAGAGCAGTGTGTGAGGGGCGAGAGAAATCAAGAGGCTCC 2035
Qy 1921 TTTGGAAGCAAAAGTTCATAAAAAATTTACTGCAAGTACTTTACCTGTTCTTCTCATGACCG 1980
Db 2036 TTTGGAAGCAAAAGTTCATAAAAAATTTACTGCAAGTACTTTACCTGTTCTTCTCATGACCG 2095
Qy 1981 GGTCTATTCTCTCTTGACCAACTCCCCCTTAGATTCAAGTTGTAAAGAAATCTTCTGTTAG 2040
Db 2096 GGTCTATTCTCTCTTGACCAACTCCCCCTTAGATTCAAGTTGTAAAGAAATCTTCTGTTAG 2155
Qy 2041 TGTCAACAGCCCTCTGGAGTCTCTCTCTCTACATCTGGAGGAGTATCTCTACATCCAA 2100
Db 2156 TGTCAACAGCCCTCTGGAGTCTCTCTCTCTACATCTGGAGGAGTATCTCTACATCCAA 2215
Qy 2101 TATGCAATGGTCACTGTTCAAGAGAGCAGCCGATTTTGCACAAGTTGCTGCAGAGATGG 2160
Db 2216 TATGCAATGGTCACTGTTCAAGAGAGCAGCCGATTTTGCACAAGTTGCTGCAGAGATGG 2275
Qy 2161 GAATTCACCAAGTCTAGGTAGCCAAAGATTACTGCACAAGCCACTGGGAAAGACACAGCAG 2220
Db 2276 GAATTCACCAAGTCTAGGTAGCCAAAGATTACTGCACAAGCCACTGGGAAAGACACAGCAG 2335
Qy 2221 TATAACTCTTGTGGGACGGAAATGTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAAAGAA 2280
Db 2336 TATAACTCTTGTGGGACGGAAATGTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAAAGAA 2395
Qy 2281 GGAGATAATGCACTTCTTAGATACCTGCTGGACAGGGATGATCCTTAGTGTACCTCTC 2340
Db 2396 GGAGATAATGCACTTCTTAGATACCTGCTGGACAGGGATGATCCTTAGTGTACCTCTC 2455
Qy 2341 TAAAGAACTACAGCCCCAAGTGGAAAGAGTGGACATAAATAAGTCAAGTGCACAGCTC 2400
Db 2456 TAAAGAACTACAGCCCCAAGTGGAAAGAGTGGATATAAATAAGTCAAGTGCACAGCTC 2515
Qy 2401 CACCATCTCTAGTCAAGTCAAGAGAGAAAGCCCTAAATAAAGACAGACAGAGTGAAGA 2460
Db 2516 CACCATCTCTAGTCAAGTCAAGAGAGAAAGCCCTAAATAAAGTGAAGACAGAGAGTGAAGA 2575

DE Murine pCIP gene (ortholog of human AIB1 gene).
KW AIB1; amplified in breast cancer; cancer; steroid; receptor;
KW coactivator; SCR; estrogen; ER; estrogen dependent transcription;
KW breast cancer; lung cancer; colon cancer; prostate cancer;
KW melanoma; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT 110..4453
FT /*tag= a
FT /product= pCIP_protein
PN W09857982-A2.
PD 23-DEC-1998.
PE 17-JUN-1998; U12689.
PR 17-JUN-1997; US-049728.
PA (USGO) US GOVERNMENT.
PI Meltzer P, Trent JM;
DR WPI: 99-080946/07.
DR P-PSDB: W81029.
New isolated steroid receptor co-activator, AIB1 - used to develop products for the diagnosis and treatment of steroid-responsive tumours, e.g. breast, lung, prostate or colon cancers or melanomas
PS Disclosure: Page 35-39: 57pp; English.
CC The AIB1 protein is a member of the steroid receptor coactivator-1 (SCR-1) family of nuclear receptor co-activators that interact with estrogen receptors (ER) to enhance ER-dependent transcription.
CC The AIB1 gene is amplified and over-expressed in certain cancers
CC In particular breast cancer and steroid hormone responsive cancers.
CC The AIB1 polypeptide can be used to identify compounds which inhibit ER-dependent transcription. Increased expression of the AIB1 gene indicates aberrantly proliferating cells, thus detection of increased expression of the AIB1 gene or an increase in the number of copies of the AIB1 gene can be used to diagnose cancer or a predisposition towards developing cancer. Compounds which inhibit expression of AIB1 or compounds which inhibit interaction of AIB1 with steroid receptors or nuclear co-factors can be used for reducing the proliferation of cancer cells. This sequence is of the murine pCIP gene, a mouse ortholog of AIB1.
SQ Sequence 4621 BP; 1227 A; 1244 C; 1194 G; 956 T;

Query Match 62.5%; Score 2812.2; DB 1; Length 4621;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 3547; Conservative 0; Mismatches 718; Indels 184; Gaps 12;
QY 1 GCTGGATGCTGGACTCAGAGACCAATAAATAAAGTCTTGAACATCCTTTGACTGGTT 60
26 GCTGACAGTGGACTCCGAGATCGGTAAACCAACTTTCCTGCCCTTCTCGAACAGCT 85
61 AGCCAGTGTCTGATGATATTCAGATGAGTGGATTAGGAGAAAC---TTGGATCCACT 117
86 -GTCAGTGTCTGATCTGTATCAGGATGAGTGGACTAGCGGAAGCTCTTTGGATCCGCT 144
QY 118 GGCAGTGTATTCAGAAACGCAATATTCGATGATGATGATGATGATGATGATGATG 177
145 GGCCTGTAGTCTCGGAACGCAATATTCGATGATGATGATGATGATGATGATGATG 204
QY 178 CAGTGTGTGAAACGAGAGCGGAGCAGGAGCAAGTAAATATTTGAAGATTCGCTGAGCT 237
205 CAGTGTGTGAGAGTGGCAGCGGAGCGGAGCAGGAGCAAGTAAATATTTGAAGATTCGCTGAGCT 264
QY 238 GATATCTGCCAATCTTATGATATTCAGATGATGATGATGATGATGATGATGATGATG 297
265 CATCTCTGCAATCTCAGCGATCAGCAACTTCATGTCAGCCAGATAAATGTCCTAT 324
QY 298 TTAAAGGAACAGTAAGACAGATAGTCAATATAAAGAGCAGGAGGAGGAGGAGGAGG 357
325 CCTAAAGGAGACAGTGTAGACAGATAGCGGCAATATAAAGAGCAGGAGGAGGAGGAGG 384
QY 358 TGATGATGATGTTCAAAAACGCGATGATCTTCTACAGGCGGAGGAGGAGGAGGAGGAG 417
385 TGATGATGATGTTCAAAAACGCGATGATCTTCTACAGGCGGAGGAGGAGGAGGAGGAG 444
QY 418 CUCCTTTAGACCGCTTTTACTTCAGGCATTTGGATGTTTCTTATTTGTTGGTGAATCGAGA 477

Db 445 CTCCTTAGGACCGCTTTTACTACAGGCACCTGGATGTTCTCTGTTGTTGGTGAATCGAGA 504
QY 478 GCAAAACATTTGATTTGATATCAGAAAATGTACAAATACCTGCAATATTAAGCAAGAGA 537
Db 505 TGGAAACATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 564
QY 538 CCGTGTAAACACAAGTGTTCACAAATATCTTACATCAAGAAGACAGAAAAGGATTTCTTAA 597
Db 565 CCGTGTAAACACAAGTGTTCACAAATATCTTACATCAAGAAGACAGAAAAGGATTTCTTAA 624
QY 598 GAATTTACCAAAATCTACAGTTAATGGAGTTTCTTGGACAAATGAGCCCCCAAGACAAA 657
Db 625 ACACCTTACCAAAATCCACAGTTAATGGAGTTTCTTGGACTAATGAGAACCCAGACAAA 684
QY 658 AAGCCATACATTTAATTTGCGGTATGTTGATGAAAACACACATGATATTTCTGGAAGACAT 717
Db 685 AAGCCATACATTTAATTTGCGGTATGTTGATGAAA---CACACGACATTTTGGAAAGCGT 741
QY 718 AAACGCCAGTCTGAAATGCGCCAGAGATATGAAACAATGAGTGTCTTGGCCTGCTCA 777
Db 742 GATGCCAGTCCGAAACACGCCAGAGATATGAAACAATGAGTGTCTTGGCCTGCTCA 801
QY 778 GGCAGAGCTATGATGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 837
Db 802 GCCTCGCCTATGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 861
QY 838 CGGCATTTACTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 897
Db 862 CGCGCTGACTG-----GCCATTCCCATCCAGTCTCTGAGAGCTTTATTACAGACA 912
QY 898 TGATCTTTTACGAAAGGTTGTCAATATAGATACAAATTCAGTGTGATGATGATGATGATG 957
Db 913 TGACCTTTCCGGAAGGTTGTCAATATAGATACAAATTCAGTGTGATGATGATGATGATG 972
QY 958 TGGCTTTGAAGATATAATCCGAAGGTGATTACAGAGATTTTCTAGTCTAAATGATGGCA 1017
Db 973 TGGCTTTGAAGATATAATCCGAAGGTGATTACAGAGGTTCCTCAGTCTGAATGATGGCA 1032
QY 1018 GTCATGTTCCAGAAACGCTACTATCAAGAGCTTATCTTAATGGCCATGCAGAAACCCC 1077
Db 1033 GTCATGTTCCAGAAACGCTACTATCAAGAGGTTCCTCAGTCTGAATGATGGCA 1092
QY 1078 AGTATATGATTTCTGTTGGCTGATGGAATATAGTGTGATGATGATGATGATGATGATG 1137
Db 1093 CGTGTATGTTCTGTTGGCTGATGGAATATAGTGTGATGATGATGATGATGATGATG 1152
QY 1138 CTTCCGAAATCCTGTAAACAATGATCGACATGGCTTTTCTCAACCCACTTCTTTCAGAG 1197
Db 1153 CTTCCGAAATCCTGTAAACAATGATCGTACGGCTTCATCTCGACCCACTTCTTTCAGAG 1212
QY 1198 AGAACAGATGATATAGCAACCAACCCAAATCCTGTTGGACAAGGATTTAGACACCTAT 1257
Db 1213 AGAACAGATGATATAGCAACCAACCCAAATCCTGTTGGACAAGGATTTAGACACCTAT 1272
QY 1258 GCGTGGATGCAACAGTTGCGTAGGCGGATGATGTCGCAACCAAGGCTTACAGAT 1317
Db 1273 AGCAGGT-----GTGGCGTGAGCATGTCTCCAAATCAGATGTACAGAT 1317
QY 1318 GCGAGCAGCAGCGGCTTATGGCTTGGCAGAGCCTTAGCACCACAGCGGAGATGAGTGGAGC 1377
Db 1318 GATGGCAGCGGAGCTTATGGCTGCCAGACCCACAGCAACACAGGCGAGATGGGTGGAGC 1377
QY 1378 TAGGTATGGGGTTCAGTACATAGCTTATTTACCCCTTGGCCAGGCGATGCAATACCC 1437
Db 1378 TAGGTATGGGGTTCAGTACATAGCTTATTTACCCCTTGGCCAGGCGATGCAATACCC 1437
QY 1438 ATCTTCTTACAGAACAACTATGGGTCAACTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1497
Db 1438 ATCTTCTTACAGAACAACTATGGGTCAACTAGTGTAGTGTAGTGTAGTGTAGTGTAGT 1497
QY 1498 TGGCTTCCCCCAACACAGCAGATATCATGATTTCTCTCGTAATCGTGGGAGTCCAAA 1557

Db 3048 AACAGCAGCAGCAACAGCAGCAGATGCTTCAATGAGAACTGGTGAGATTCCTCCATGGAA 3107
Qy 3063 TGGGGCTAATCCCTATGCGCAAGCAGCAGCATCTAACCAACTGGTTCCTGGCCCGATG 3122
Db 3108 TGGGAGTCAATCCNTATAGCCAGCAGTGCAGTCTAACCAACAGGTTCCTGGCCAGAG 3167
Qy 3123 GCATGTTGTCATGGAACAAGTTTCTCATGGCACTCAAAATAGGCTCTCTCTTAGGAAT 3182
Db 3168 GCATGCTCTATGGAACAAGTCTCAGGGTCTCAAAATAGGCTCTCTCTTAGAACT 3227
Qy 3183 CCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCAGAGTGCAGAAAGCAT 3242
Db 3228 CTCCTGGATGATCTGCTTGGGCCACCTTCTAACCGCAGAGGGCCAGAGTGCAGAGAGCTC 3287
Qy 3243 TATTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGACAGGCTCGAAGAAATG 3302
Db 3288 TGCTGGACAGCTGCACACTCTTCTGAGCAACACAGATGCGACAGGCTCGAGGAGATCG 3347
3303 ACAGAGCTTTGGGCAATTCCTGAACCTTGTCAATCAGGGGACAGCATTAGAGCCCAACAGG 3362
3348 ACAGGSCCTTGGGAATTCCTGAGCTCGTAATCAGGGACAGCTTTGGAGTCCAAACAGG 3407
Qy 3363 ATGCTTTCCAGGCGCAAGCAGCAGTAATGATGGATCAGAGCAGGATATATATGGAC 3422
Db 3408 ATGTTTTCAGAGGCCAAGCAGCAGTAATGATGGATCAGAGGCTGCACATATATGGAC 3467
Qy 3423 AGACATACCCAGCAGCAGGCTTCAATGCAAGGAGGCTTTCATCTTCAGGACATAC 3482
Db 3468 AGACATACCCAGCTCAGGCTCTCCCTTCAAGGAGGCTTTAACTTCAGGACAGTAC 3527
Qy 3483 CATCTTTAACTCTATGATGAATCAGATGAACACAGCAGGCAATTTCTCTCCAAGAA 3542
Db 3528 CATCTTTAACTCTATGATGGTCAAGTATGCGCAGCAGGAGCTTCTCTGCAAGCA 3587
Qy 3543 TGCACCCAGCAGCAACATCATGAGACCCCGCAACACACCCCAAGCAACTTAGATGC 3602
Db 3588 TGCATCTTAGAGCCGCTCGTGAGACCAAGGACCAACACCCCAAGCAGCTGAGATGC 3647
Qy 3603 AGCTTCAGCAGGCTCAGGCGCAGCAGTTTGAATCAGAGCCGACAGCAGCTTGAT 3662
Db 3648 AGCTTCAGCAGGCTCAGGCGCAGCAGTTTGAATCAGAGCCGACAGCAGCTTGAA 3707
Qy 3663 TGAATGGAACCCCTACTGCTGTGTGCTGCTGAGGCTGATGAGGCTATGATGAGCC 3722
Db 3708 TGAATGGAACCC-----CTGCTGCACTGCTGTGATGAGCCCATGATG-----C 3755
3723 AGCAGGTTTCTTAATGCTCAATGCTGCGCCCAACGACGAGAGCTGTGATGATCATC 3782
3756 CCCAGGCTTCTTAAATGCCAAATGGCTGCGCCAGCAGAAACGAGAGCTGATGAGCCATC 3815
Qy 3783 ACTTCGACACAGAGGCTGCTATGATGATGATGATGATGATGATGATGATGATGATG 3842
Db 3816 ACCTGCACAGCAGAGGATGGCGATGATGATGATGATGATGATGATGATGATGATGAT 3847
3843 AGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAACAGCAGCAACAGCAGCAGCAGCAGC 3902
Db 3847 -----GTCCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3875
Qy 3903 CACCTCTTAATGTGACTGCTTCCCGCCAGCATGGATGGGCTTTTGGCAGGACCCCAATGC 3962
Db 3876 CACCTCCCAAGCTCACCGCTCCCGCCAGCATGGAGGGGTTTGGCAGGTTTCAGCAATGC 3935
Qy 3963 CACAGCTCTCCGCAACAGTTTCCATATCAACCAAAATTTGGAATGGGACCAACACAG 4022
Db 3936 CGCAAGCCCTCCCAACAGTTTCCATATCCAGCAAAATTTAGCAACGGGACCAACACAG 3995
Qy 4023 -ATCCAGCTTTGGTGGAGTGTCTAGTCTCTCCCAATGCAATGATGCTCAAGATGGT 4081
Db 3996 TAGCCAGCTTTGGTGGAGTGTCTAGTCTCTCCAGTGCATGATGATGATGATGATGATG 4055
Qy 4082 CCCTCCCAAGATCCCATGATGCAACACCCGAGGCTGCATCTATCATCTATCATCTCGATG 4141
Db , 4056, COTCCCAAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4115

Qy 4142 ATGAAGGCTGGCCATCAGGAATTTGGCCAGGAACAGCTCTCTTTTCCAGCAGCAGTTT 4201
Db 4116 ATGAAGGCTGGCCATCAGGAACCTGGCCAGGAATGGCTCTTCCCGCCAGCAGTTT 4175
Qy 4202 GCCCACCAGGGAATCTCGAGTGTATAGTATGTGTGCACATGAATGCGCAGTGGTCAC 4261
Db 4176 GCTCCCGAGGGAACCTCGCAGCTACAACATGTGTATGAACAGCAGCGGTGGGCAC 4235
Qy 4262 ATGGCAGAGTGAACATGAACCCCATGCCCATGTCTGGCATGTATGGTCTGTATCAG 4321
Db 4236 TTGGCAGAGTGGCAGTGAACCCCATGCCCATGTCTGGCATGTATGGTCTGTATCAG 4295
Qy 4322 AAATACGTGACATCTGACACAGGACCTCTTAAAGGAACCACTGTACAAATGACAT 4381
Db 4296 AAATACGTGACATCTGACATGCTAGTGGGAC-----TGACTGTACATGACATGACAT 4348
Qy 4382 GCATAGGATTTATGGGAAGGAATCATTTTCCAGGATCATCTTGAAGAAAGGAC 4439
Db 4349 GGATCATCAGGACCTGGCGGAGCTCATTTGCTAAGCATCAGCTTGAAGCAAGGCC 4406

RESULT 4

V03517

ID V03517 standard; cDNA; 6156 Bp.

AC V03517;

DT 20-JUL-1998 (first entry)

DE Human transcriptional intermediary factor-2 (TIF2) cDNA.

KW Transcriptional intermediary factor; TIF2; human; drug screening;

KW assay; nuclear receptor; ds.

OS Homo sapiens.

PH Key

ET Location/Qualifiers

FT CDS

FT WO9802455-A2.

PD 22-JAN-1998.

PF 11-JUL-1997; US-021247.

PR 12-JUL-1996; UI-021247.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PA (CNRM) CENT NAT RECH SCI.

PA (INRM) INST NAT SANTE & RECH MEDICALE.

PA (UYPA) UNIV PASTEUR LOUIS.

PI Chambon P, Gronemeyer H, Lutz Y, Voegel J;

DR P-PSDB: W42632.

DR New isolated transcriptional intermediary factor-2 - useful in

PT assays for drugs which are capable of enhancing or inhibiting

PT nuclear receptor-mediated pathways

PS Claim 1; Fig 1A-B; 119pp; English.

CC This cDNA clone codes for a novel nuclear receptor transcriptional

CC mediator (see W42632) designated transcriptional intermediary

CC factor-2 (TIF2). TIF2 interacts directly with the ligand binding

CC domains (LBD) of several nuclear receptors (NR) in an agonist- and

CC AF2-integrity-dependent manner in vitro and in vivo, harbours an

CC autonomous AF, relieves NR autoquenching, and enhances the

CC activity of NR AF2s when overexpressed in mammalian cells.

CC Screening of a human placenta cDNA expression library with an

CC estradiol-bound 32P-labelled estrogen receptor probe yielded a cDNA

CC that encoded a protein fragment (TIF2.1) that interacted on

CC Far-Western blots with different 32P-labelled NR-LBDs ER, RAR, RXR

CC in an agonist-dependent manner. The TIF2 coding sequence was

CC obtained upon rescanning with a TIF2.1 cDNA probe. TIF2 cDNA has

CC been deposited as ARCC 97612. Polynucleotides encoding full-length

CC TIF2 protein or encoding TIF2 polypeptide fragments comprising

CC amino acids 624-869, 624-1131, 1010-1131 or 1288-1464 are claimed.

CC Recombinant methods for making TIF2 polypeptides are provided, as

CC are screening methods for identifying agonists and antagonists of

CC nuclear receptor AF-2 function, TIF2 AD1 activity and TIF2 AD2

CC activity. The products are useful in assays for identifying drugs

CC capable of enhancing or inhibiting NR-mediated pathways. They can

CC also be used for detection and localisation.

SQ Sequence 6156 Bp; 1799 A; 1490 C; 1406 G; 1461 T;

	Query Match	1.6%;	Score 73.2;	DB 1;	Length 168;	
	Best Local Similarity	69.7%;	Pred. No. 1.8e-10;			
	Matches	99;	Conservative	0; Mismatches	43; Indels	0; Gaps
	OY	3768	AGCTGCTAAGTCACTACTCCGACAACAGAGGGTGCGTTATGTATGCATGCAGCAGCAGCAAC	3827		
	Dg	17	AGC	76		
	OY	3828	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAACAGCAGCAACAGCACAGCAGCAGCAGCAAA	3887		
	Dg	77	AGC	136		
	OY	3888	CCCAGGCCCTTCAGCCCCACCCTCC	3909		
	Dg	137	AGCAGCACCTCAGCAGGGCTCC	158		
	RESULT	14				
	Q84834	ID	Q84834 standard; DNA; 171 BP.			
	AC	Q84834;	1995 (first entry)			
	DE	Spinocerebellar ataxia type 1 CAG repeat region patient #4.				
	KW	Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;				
	OS	CAG repeat region; patient #4; ss.				
	PN	Homo sapiens.				
	PR	WO9501437-A.				
	PF	12-JAN-1995.				
	PP	29-JUN-1994; UO7336.				
	PR	29-JUN-1993; US-084365.				
	PR	28-JUN-1994; US-267803.				
	PI	(MNU) UNIV MINNESOTA.				
	PA	Chung M, Orr HT, Zoghbi HY;				
	DR	WFI; 95-061001/08.				
	PT	New autosomal dominant spinocerebellar ataxia type 1 nucleic acid				
	PT	- used to develop prods. for detection or presymptomatic				
	PS	diagnosis of a SCAl disorder				
	PS	Disclosure; Fig 2; l1ipp; English.				
	CC	Q84831-Q84835 show the CAG repeat regions of five individuals				
	CC	affected with spinocerebellar ataxia type 1 (SCA 1). It is within				
	CC	CAG repeat region (Q84804) that the mutations responsible for				
	CC	SCA 1 occur. The full nucleic acid (Q84793) and its protein product				
	CC	(R7111l) can be used to develop products, for the presymptomatic				
	CC	detection of a SCA 1 disorder.				
	SQ	Sequence	171 BP; 52 A; 59 C; 56 G; 4 T;			
	Query Match	1.6%;	Score 73.2;	DB 1;	Length 171;	
	Best Local Similarity	69.7%;	Pred. No. 1.8e-10;			
	Matches	99;	Conservative	0; Mismatches	43; Indels	0; Gaps
	OY	3768	AGCTGCTAAGTCACTACTCCGACAACAGAGGGTGCGTTATGTATGCATGCAGCAGCAGCAAC	3827		
	Dg	20	AGC	79		
	OY	3828	AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAAACAGCAACAGCAGCAACAGCACAGCAGCAGCAAA	3887		
	Dg	80	AGC	139		
	OY	3888	CCCAGGCCCTTCAGCCCCACCCTCC	3909		
	Dg	140	AGCAGCACCTCAGCAGGGCTCC	161		
	RESULT	15				
	T42476/c	ID	T42476 standard; DNA; 645 BP.			
	AC	T42476;				
	DT	01-AUG-1997 (first entry)				
	DE	Hunt14 coding sequence.				
	KW	Hunt14; Huntington's disease; ITL5; Huntington gene; gene therapy;				
	KW	Huntington's disease; HD; autosomal dominant; neurodegenerative disease;				

This Page Blank (uspto)

WIPSAHA

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

on: Fri Sep 17 21:02:33 1999; MasPar time 48.85 Seconds
Tabular output not generated. 616.024 Million cell updates/sec

Title: >US-09-041-994-2

Description: (1-1415) from US09041994.pap

Perfect Score: 9849
Sequence: 1 MSGLENDPLASDSKRRL.....MNNMPMSGMPGPDQKYC 1415

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35

1:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 39.485; Variance 214.637; scale 0.184

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	3346	34.0	1464	30	Human transcriptional	2.04e-248
2	914	9.3	147	34	Clone AJL1.1 protein s	1.24e-57
3	643	6.5	1061	24	Human steroid recepto	6.23e-37
4	232	2.4	846	37	Human CLOCK protein.	7.96e-07
5	232	2.4	848	14	Human Ah receptor pro	7.96e-07
6	232	2.4	848	23	Mouse CLOCK protein.	7.96e-07
7	235	2.4	855	37	Human Ah receptor.	4.96e-07
8	225	2.3	371	38	Epitope tagged TAP pr	2.39e-06
9	231	2.3	805	14	Murine Ah receptor pr	9.32e-07
10	213	2.2	1313	33	Spinocherebellar ataxi	1.56e-05
11	202	2.1	539	30	Yeast transcriptional	8.55e-05
12	209	2.1	594	34	Mouse neuronal PAS do	2.90e-05
13	210	2.1	800	14	Transcription factor	2.48e-05
14	202	2.1	870	27	Human endothelial PAS	8.55e-05
15	203	2.1	914	25	Spinocherebellar ataxi	7.33e-05
16	203	2.1	1312	30	Human ataxin-2.	7.33e-05

17	210	2.1	2414	30	W40057	Cellular transcriptio	2.48e-05
18	210	2.1	2414	14	R84882	Transcription factor	2.48e-05
19	205	2.1	2441	30	W40058	Cellular transcriptio	5.38e-05
20	205	2.1	2441	14	R79054	CREB binding protein.	5.38e-05
21	204	2.1	3144	11	R58777	Protein encoded by Hu	6.28e-05
22	204	2.1	3144	29	W44742	Human huntingtin prot	6.28e-05
23	204	2.1	3144	22	W09871	Human huntingtin.	6.28e-05
24	204	2.1	3144	26	W36887	Previously undescribe	6.28e-05
25	193	2.0	306	14	R75454	Polyoma virus enhance	3.39e-04
26	193	2.0	513	14	R75453	Polyoma virus enhance	3.39e-04
27	194	2.0	548	39	W89189	Alternatively spliced	2.91e-04
28	199	2.0	590	34	W58091	Human neuronal PAS do	1.36e-04
29	194	2.0	596	39	W39184	Osfl/Cbfa1 native pol	2.91e-04
30	195	2.0	657	25	W30758	Murine SIM (single-mi	2.50e-04
31	197	2.0	737	15	R76639	Deltex protein produc	1.84e-04
32	197	2.0	737	15	R76640	Deltex protein.	1.84e-04
33	197	2.0	737	23	W18317	Drosophila Deltex pro	1.84e-04
34	200	2.0	875	27	W37098	Murine endothelial PA	1.16e-04
35	190	1.9	81	30	W41372	SCA2 protein fragment	5.36e-04
36	189	1.9	373	20	W06559	HIF-1 alpha delta-NB	6.24e-04
37	184	1.9	591	32	W48796	Drosophila melanogaste	1.33e-03
38	184	1.9	609	37	W83215	Human h-NUMB-R.	1.33e-03
39	183	1.9	678	8	R42087	D. melanogaster dorsa	1.55e-03
40	189	1.9	789	29	W39927	Human Arnt receptor p	6.24e-04
41	189	1.9	805	20	W06558	Hypoxia inducible fac	6.24e-04
42	190	1.9	816	34	W68094	Mouse neuronal PAS do	5.36e-04
43	190	1.9	816	13	R71111	Spinocherebellar ataxi	5.36e-04
44	189	1.9	826	36	W80418	Amino acid sequence o	6.24e-04
45	189	1.9	826	20	W06557	Human hypoxia inducib	6.24e-04

ALIGNMENTS

RESULT	ID	W42632 standard; Protein; 1464 AA.
AC	W42632	20-JUL-1998 (first entry)
DE	Human transcriptional intermediary factor-2 (TIF2).	
KW	Transcriptional intermediary factor; TIF2; human; drug screening;	
OS	Homo sapiens.	
PH	Key	Location/Qualifiers
FT	Peptide	13..20
FT	Peptide	/note= "nuclear localisation signal"
FT	Peptide	31
FT	Domain	/note= "nuclear localisation signal"
FT	Domain	1010..1131
FT	Domain	/note= "activation domain AD1"
FT	Domain	1288..1464
FT	Protein	/note= "activation domain AD2"
FT	Protein	624..1287
FT	Protein	/note= "cytoplasmic TIF2.1 polypeptide, functional coactivator fragment"
FT	Protein	624..869
FT	Protein	/note= "cytoplasmic TIF2.5 polypeptide, interacts with NID domain of nuclear receptors, but does not enhance transcription"
FT	Protein	1010..1179
FT	Protein	/note= "cytoplasmic TIF2.8 polypeptide, enhances transcription but does not bind nuclear receptor"
FT	Protein	940..1131
FT	Protein	/note= "cytoplasmic TIF2.12 polypeptide, enhances transcription but does not bind nuclear receptor"
FT	Protein	1288..1464
FT	Protein	/note= "TIF2.2 polypeptide, enhances CBP-independent transcription"
FT	Region	865..1061
FT	Region	/note= "dnSRC-1, dominant negative fragment"
PN	W09802455-A2.	
PD	22-JAN-1998.	
PF	11-JUL-1997; U12100.	

PN W09825962-A2.
 PD 18-JUN-1998.
 PF 12-DEC-1997; U23224.
 PR 11-DEC-1997; US-989232.
 PR 13-DEC-1996; US-766263.
 PA (GEM) J GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoey JM, Merberg D,
 PI Racine LA, Spaulding V, Treacy M;
 DR WPI; 98-362424/31.
 DR N-PSDB; V44857.
 PT New isolated polynucleotides - obtained from human adult testis,
 PT human adult ovary, human adult brain and human adult heart cDNA
 PT libraries
 PS Claim 8; Page 62-63; 108pp; English.
 CC This sequence is secreted protein, encoded by a polynucleotide of the
 CC invention. The DNA was isolated from a human adult testis cDNA library,
 CC and is designated clone A11.1. The DNA sequences and encoded
 CC polypeptides can be used as nutritional sources or supplements, or may
 CC exhibit e.g. cytokine and cell proliferation/differentiation activity,
 CC immune stimulating or suppressing activity, haematopoiesis regulating
 CC activity, receptor/ligand activity, anti-inflammatory activity,
 CC activin/inhibin activity, chemostatic/chemokinetic activity,
 CC cadherin/tumour invasion suppressor activity, tissue growth activity,
 CC tumour inhibition activity or other activities.
 SQ Sequence 147 AA;

Query Match 9.3%; Score 914; DB 34; Length 147;
 Best Local Similarity 100.0%; Pred. No. 1.24e-57;
 Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 msqctstipsssqekdkkttseegsgdldaldilgdtssdfynnsisangshlg 60
 QY 766 MSQCTSTIPSSSQEKDPKIKTSESGDLNLDAILGDTSSDFYNNSSISNGSHLG 825

Db 61 tkqvgfgtnslgkssgsvgsirppynravslgsvsgsppvknlsafmpkqpm 120
 QY 826 TKQVFGTNSLGLKSSQSVGSIRPPYNRAVSLDSPVSGSSPPVKNLSAFPMPKQPM 885

Db 121 ggnprmdsgenyssmg 138
 QY 886 GGNPRMDSQENYSSMG 903

RESULT 3
 ID W26370 standard; Protein; 1061 AA.
 AC W26370;
 DT 04-DEC-1997 (first entry)
 DE Human steroid receptor coactivator-1 (SRC-1).
 KW Steroid receptor coactivator-1; SRC-1; molecular switch;
 KW gene therapy; transgenic animal.
 OS Homo sapiens.
 PN W09710337-A1.
 PD 20-MAR-1997.
 PF 20-AUG-1996; U13482.
 PR 15-SEP-1995; US-003784.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PI O'Malley BO, Onate S, Tsai M, Tsai SY;
 DR WPI; 97-202233/18.
 DR N-PSDB; T84543.
 PT Steroid receptor coactivator-1 polypeptide and DNA - regulated using
 PT a molecular switch, used in gene therapy
 PS Claim 3; Fig 1; 116pp; English.
 CC This polypeptide comprises human steroid receptor coactivator-1
 CC (SRC-1), a novel protein required for human progesterone receptor
 CC (hPR) transactivational function. SRC-1 acts as a coactivator for
 CC hPR by reversing receptor squelching. It is a coactivator for
 CC multiple steroid receptor superfamily members. The truncated
 CC C-terminal region of SRC-1 acts as a dominant negative regulator
 CC of steroid receptor function. An SRC-1 related disease can be
 CC treated by introducing SRC-1 nucleic acid (see T84543) into a host
 CC cell and infusing the cells into the patient causing an increase in
 CC the transcription of SRC-1 (claimed). A molecular switch can be
 CC used to regulate expression of a nucleic acid cassette incorporating

CC an SRC-1 coding region for use in gene therapy. Transcription of a
 CC target gene can be decreased by providing a nucleic acid encoding a
 CC dominant-negative inhibitor of a SRC-1 polypeptide in a cell
 CC containing the target gene (also claimed).
 SQ Sequence 1061 AA;

Query Match 6.5%; Score 643; DB 24; Length 1061;
 Best Local Similarity 36.2%; Pred. No. 6.23e-37;
 Matches 113; Conservative 89; Mismatches 89; Indels 21; Gaps 18;

Db 456 mdgavtsvkseilpaslqsatrptsrln-rilpeleleaidnqfgpgtgddai-pwtin 513
 QY 954 LGSIPILPLRSNIPGA-RVLIQQQQOMLQMRPGEIPLMGANGPYGAAASNOLGSWPD 1012

Db 514 nvtainqsk-sedqci-sseqldellcppttvegrndekalleivslgskdetelael 571
 QY 1013 GMLSMQSVSHGTQNRPLRLNSLDLVGPPSNLEQSDERALLDQLHTLLSNTDGLLEEI 1072

Db 572 dralgidklv-qggglavlserfppqatpplmeerpnlhysqyspfpftanlpsfqg 630
 QY 1073 DRALGIPELYNQGALEPKQDAFQGEAA--VMDQKAGLYGQTYPAQGP-P-MQGGFH- 1127

Db 631 mvrqkpslgtmpvqvtprgafspgmqprqtlnrppa-apnqlrlqlqrlcgqgqli 689
 QY 1128 LQGSPSFNSMMNQ--QGNF-PLQGHPRANTRPTNPKOLRMQLQQLQGGQ--FL 1184

Db 690 hqnrgai-lhqfaatpvginrmgmcqqitpqpplnaqlagrqrelysqhqrq-r-ql 746
 QY 1185 NQSRQALELKMENPTAGGAAMRPMQ--QGFLLNAQVMAQSRRELLSHHFQQRVAM 1241

Db 747 lqqqramlrrq 758
 QY 1242 MMQQQQQQQQQQQ 1253

RESULT 4
 ID W79533 standard; Protein; 846 AA.
 AC W79533;
 DT 02-FEB-1999 (first entry)
 DE Human CLOCK protein.
 KW CLOCK; circadian rhythm; human; jet-lag; sleep-wake disorder;
 KW seasonal affective disorder; cancer; transcription factor;
 KW diagnosis; therapy.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Protein 2..846
 FT /note= "Claimed fragment"
 FT Protein 10..846
 FT /note= "Claimed fragment"
 FT Protein 11..846
 FT /note= "Claimed fragment"
 FT Domain 35..81
 FT /label= bHLH
 FT /note= "basic helix-loop helix domain"
 FT Domain 115..163
 FT /label= PAS-A
 FT Domain 273..318
 FT /label= PAS-B
 PN W09840514-A1.
 PD 17-SEP-1998.
 PF 13-MAR-1998; U05114.
 PR 30-JUN-1997; US-885291.
 PR 13-MAR-1997; US-816693.
 PA (NOUN) UNIV NORTHWESTERN
 PI Pinto LH, Takahashi JS, Turek F;
 DR WPI; 98-520828/44.
 DR N-PSDB; V61450.
 PT Novel mouse and human circadian rhythm gene, clock - useful for
 PT treating e.g. jet-lag, sleep-wake disorders, abnormal cell division,
 PT etc
 PS Claim 20; Fig 14.1-14.3; 154pp; English.
 CC This is the amino acid sequence of human CLOCK protein, an integral
 CC component of the circadian clock that serves to regulate various

LD	W25000 standard, protein, 846 aa.
AC	W25668;
DT	05-NOV-1997 (first entry)
DE	Human Ah-receptor.
KW	Human; Ah receptor; yeast; nuclear translocator; reporter gene;
KW	dioxin responsive element; environmental pollutant.
OS	Homo sapiens.
FH	Key
FT	Location/Qualifiers
FT	13..41
FT	/note= "Basic region"
FT	43..84
FT	/note= "Helix-loop-helix domain"
FT	113..400
FT	/note= "DBAS domain"
FT	

[illegible]

WPI; 95-051315/07.
N-PSDB; Q99601.
Nucleic acid sequences encoding murine and human Ah receptors -
used in competitive binding assays to detect environmental
pollutants
Claim 1; Column 31-36; 24pp; English.
The amino acid sequence of the murine Ah receptor protein. The Ah
receptor is a soluble protein which mediates an individual's response to a
variety of drugs, carcinogens and toxic agents. The gene was isolated
from a mouse genomic DNA library using the probes Q99603-5. These probes
were determined from the N-terminal sequence of a purified Ah receptor,
which was purified from C57BL/6J mouse liver. The screen isolated a
clone designated cAh1 containing the full length cDNA. The 1.4 kb
fragment of the murine clone was used to obtain clones covering the
coding sequence of the human AH receptor (Q99602). The Ah receptors are
useful in bioassays to detect environmental pollutants. The genes can be
used to generate recombinant organisms useful as biomonitors for
environmental pollutants.
Sequence 805 AA;

Query Match 2.3%; Score 231; DB 14; Length 805;
Best Local Similarity 30.3%; Pred.No. 9.32e-07;
Matches 46; Conservative 46; Mismatches 52; Indels 8; Gaps 6;

D Db 26 paeigknspskrhrrlnrteidrtasllp--fpq-dvin-kldklsvlrtsvylraksf 81
::: : |||::: : |||::: : |||::: : |||::: : |||::: : |||::: : |||:
Q Y 25 PGOGLTCSEKRRRQESKYIEELAEILSANLSIDINFNKPKCAILKETVQIRQIKE 84
D Db 82 fdvalkstpadrngggdqcrag-lrdwdqlgegefllqalngvlvtadalfvyasstl 140
::: : |||::: : |||::: : |||::: : |||::: : |||::: : |||::: : |||:
Q Y 85 OGKTI-SNDDDYQKADVSSTGGVIDKDLSL-GPLLALGDGFLFVNREANIVFSENV 141
141 qdylfqsqdvihgsyvelihtedraefqrql 172
::: : |||::: : |||::: : |||::: : |||::: : |||::: : |||::: : |||:
Q Y 142 TQYLQKQEDLVNTSVYNILHEEDRKDFKNL 173

RESULT 10
ID W60213 standard; Protein; 1313 AA.
AC W60213;
DE 02-OCT-1998 (first entry)
DT Spinocerebellar ataxia type 2 (SCA2) disease associated protein.
DE Spinocerebellar ataxia type 2; SCA2; gene therapy; antisense therapy;
KW CAG repeat; neurodegenerative disease.
OS Homo sapiens.
FT Key Location/Qualifiers
FH Region 166..188
FF /note= "Glutamine rich region; this region is increased in SCA2 patients"

W0918920-A1.
PD 07-MAY-1998.
PF 30-OCT-1997; J03946.
PR 30-OCT-1996; JP-304059.
PA (SRLS-) SRL INC.
PI Sanpei K, Tsuji S;
DI WPI: 98-272215/24.
DR N-PSDB; V30270.
PT Nucleic acid fragments associated with spinocerebellar ataxia type 2
FT - contain increased number of CAG repeat region compared to normal
PT gene
PS Claim 1; Pages 13-22; 38pp; Japanese.
CC This is the protein sequence of a gene causative of spinocerebellar
CC ataxia type 2 (SCA2), a neurodegenerative disease. The gene associated
CC with SCA2 has a tri-nucleotide (CAG) repeat region which in the
CC expression product produces a polyglutamine sequence from Gln-166 to
CC Gln-188. In the normal gene there are 15-25 CAG repeats but in SCA2
CC patients this number is increased to 35-100. Peptides encoded by nucleic
CC acid fragments (DNA or RNA) containing sequences from the SCA2 associated
CC gene, antibodies recognising the peptides and antisense nucleic acids
CC hybridising with the nucleic acid fragments can be used for the
CC investigation and diagnosis of SCA2. They can also be used for the
CC treatment of SCA2 by antisense therapy or gene therapy.
SQ -sequence 1313 AA;

Search completed: Fri Sep 17 21:06:34 1999
Job time : 241 secs.

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
on: Fri Sep 17 21:16:43 1999; MasPar time 16.36 Seconds
Tabular output not generated. 878.195 Million cell updates/sec

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pep
Sequence: 1 MSGLGENDPLASDSRKRL.....MNNPMPMSGMPGPDQKYC 1415

Scoring table: PAM 150
Gap 11

Searched: 106580 seqs, 10152877 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PC9_COMB 4:backfiles1

Statistics: Mean 36.592; Variance 199.242; scale 0.184

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	232	2.4	848	1	US-08-366- Sequence 4, Applicatio	1.87e-07
2	232	2.4	848	1	US-08-045- Sequence 4, Applicatio	1.87e-07
3	235	2.4	855	2	US-08-816- Sequence 2, Applicatio	1.16e-07
4	231	2.3	805	1	US-08-366- Sequence 2, Applicatio	2.19e-07
5	231	2.3	805	1	US-08-045- Sequence 2, Applicatio	2.19e-07
6	210	2.1	360	2	US-08-531- Sequence 2, Applicatio	6.09e-06
7	209	2.1	594	2	US-08-785- Sequence 6, Applicatio	7.13e-06
8	202	2.1	870	1	US-08-785- Sequence 4, Applicatio	2.13e-05
9	210	2.1	2414	3	PC9-US95-0 Sequence 2, Applicatio	6.09e-06
10	210	2.1	2414	1	US-08-227- Sequence 2, Applicatio	6.09e-06
11	205	2.1	2441	2	US-08-194- Sequence 2, Applicatio	1.34e-05
12	204	2.1	3144	1	US-08-246- Sequence 6, Applicatio	1.56e-05
13	204	2.1	3144	1	US-08-453- Sequence 6, Applicatio	1.56e-05
14	204	2.1	3144	2	US-08-457- Sequence 42, Applicati	1.56e-05
15	197	2.0	303	2	US-08-185- Sequence 5, Applicatio	4.65e-05
16	199	2.0	590	2	US-08-785- Sequence 5, Applicatio	3.41e-05
17	197	2.0	737	2	US-08-185- Sequence 4, Applicatio	4.65e-05
18	197	2.0	737	2	US-08-185- Sequence 2, Applicatio	4.65e-05
19	193	2.0	788	2	US-08-918- Sequence 4, Applicatio	8.66e-05
20	194	2.0	816	2	US-08-267- Sequence 9, Applicatio	7.41e-05
21	200	2.0	875	1	US-08-785- Sequence 3, Applicatio	2.92e-05
22	189	1.9	373	3	PC9-US96-1 Sequence 5, Applicatio	1.61e-04
23	189	1.9	373	2	US-08-480- Sequence 3, Applicatio	1.61e-04

24	183	1.9	678	3	PCT-US93-0	Sequence 3, Applicatio	4.05e-04
25	189	1.9	805	3	PCT-US96-1	Sequence 4, Applicatio	1.61e-04
26	189	1.9	805	2	US-08-480-	Sequence 4, Applicatio	1.61e-04
27	191	1.9	810	1	US-08-785-	Sequence 7, Applicatio	1.18e-04
28	190	1.9	816	2	US-08-816-	Sequence 53, Applicati	1.38e-04
29	190	1.9	816	2	US-08-785-	Sequence 8, Applicatio	1.38e-04
30	183	1.9	824	2	US-08-816-	Sequence 52, Applicati	4.05e-04
31	183	1.9	824	2	US-08-785-	Sequence 7, Applicatio	4.05e-04
32	189	1.9	826	2	US-08-480-	Sequence 2, Applicatio	1.61e-04
33	189	1.9	826	3	PCT-US96-1	Sequence 2, Applicatio	1.61e-04
34	189	1.9	826	1	US-08-785-	Sequence 6, Applicatio	1.61e-04
35	174	1.8	542	1	US-08-328-	Sequence 8, Applicatio	1.60e-03
36	174	1.8	542	1	US-07-814-	Sequence 13, Applicati	1.60e-03
37	174	1.8	542	1	US-08-258-	Sequence 13, Applicati	1.60e-03
38	177	1.8	542	3	PCT-US92-1	Sequence 63, Applicati	1.60e-03
39	177	1.8	713	1	US-08-190-	Sequence 13, Applicati	1.01e-03
40	178	1.8	747	2	US-08-816-	Sequence 51, Applicati	8.70e-04
41	174	1.8	2703	2	US-08-185-	Sequence 19, Applicati	1.60e-03
42	167	1.7	428	1	US-08-190-	Sequence 29, Applicati	1.60e-03
43	163	1.7	521	3	PCT-US93-0	Sequence 10, Applicati	8.41e-03
44	165	1.7	795	1	US-07-716-	Sequence 5, Applicatio	6.23e-03
45	160	1.6	760	1	US-08-195-	Sequence 2, Applicatio	1.32e-02

ALIGNMENTS

RESULT 1
ID US-08-366-051B-4 STANDARD; PRT; 848 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 4, Application US/08366051B
XX
Sequence 4, Application US/08366051B
CC Patent No. 5650283
CC GENERAL INFORMATION:
CC APPLICANT: Bradfield, Christopher A.
CC APPLICANT: Dolwick, Kristin M.
CC APPLICANT: Carver, Lucy A.
CC TITLE OF INVENTION: An Receptor cDNAs and Genetically
CC TITLE OF INVENTION: Engineered Cells for Detecting Agonists to the Ah
CC TITLE OF INVENTION: Receptor
CC NUMBER OF SEQUENCES: 36
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Tilton, Fallon, Lungmus & Chestnut
CC STREET: 100 South Wacker Drive, Suite 960
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606-4002
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/366,051B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Tilton, Timothy L.
CC REGISTRATION NUMBER: 16,926
CC REFERENCE/DOCKET NUMBER: NU-9207-CIP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312)-456-8000
CC TELEFAX: (312)-456-7776
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 848 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

CC	STREET:	268 BUSH STREET, SUITE 3200
CC	CITY:	SAN FRANCISCO
CC	STATE:	CALIFORNIA
CC	COUNTRY:	USA
CC	ZIP:	94104
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk
CC	COMPUTER:	IBM PC compatible
CC	OPERATING SYSTEM:	PC-DOS/MS-DOS
CC	SOFTWARE:	PatentIn Release #1.0, Version #1.30
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/785.310A
CC	FILING DATE:	21-JAN-1997
CC	CLASSIFICATION:	536
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME:	OSMAN, RICHARD A.
CC	REGISTRATION NUMBER:	36,627
CC	REFERENCE/DOCKET NUMBER:	UTSD:1226
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	(415) 343-4341
CC	TELEFAX:	(415) 343-4342
CC	INFORMATION FOR SEQ ID NO:	6:
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	594 amino acids
CC	TYPE:	amino acid
CC	STRANDEDNESS:	single
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	peptide
CC	SEQUENCE	594 AA; 63/36 MW; 1871071 CN;
SQ	Query Match	2.1%; Score 209; DB 2; Length 594;
	Best Local Similarity	28.5%; Pred. No. 7.13e-06;
	Matches	35; Conservative 35; Mismatches 50; Indels 3; Gaps
Db	123	RRGPVALVE-VFEQH-LGHHILQSLDGVFVALNQEGKELYISVTYSIYGLSQVELTGS 180
QY	96	QKADVSTGGQVIDKSLGLPLLIQLALDGLFVFNREANIVFVSENVTQYLYKKQEDLVNT 155
Db	181	SVFYDTHPGDHEVLEFQLGLRAAS-IGPTTPSVSSSSSSSSSSSLVDTEIEASPTASP 239
QY	156	SVNIUHEEDRKDFLKNLPKTVNGSVWTEPQOKSHTFNCRMLMKTPHDILEDINASP 215
Db	240	AFR 242
QY	216	EMR 218
RESULT	8	
ID	US-08-785-241-4	STANDARD; PRT: 870 AA.
XX	xxxxxx	
XX		
DT		
XX		
DE	Sequence 4, Application US/08785241	
XX		
CC	Sequence 4, Application US/08785241	
CC	Patent No. 565963	
CC	GENERAL INFORMATION:	
CC	APPLICANT:	McKnight, Steven L.
CC	APPLICANT:	Russell, David W.
CC	APPLICANT:	Tian, Hui
CC	TITLE OF INVENTION:	Endothelial PAS Domain Protein
CC	NUMBER OF SEQUENCES:	7
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESS:	SCIENCE & TECHNOLOGY LAW GROUP
CC	STREET:	268 BUSH STREET, SUITE 3200
CC	CITY:	SAN FRANCISCO
CC	STATE:	CALIFORNIA
CC	COUNTRY:	USA
CC	ZIP:	94104
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Floppy disk

Sequence 2, Application US/08194468
Sequence 2, Application US/08194468
Patent No. 5750336
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
TITLE OF INVENTION: RESPONSIVE GENES
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
City: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,468
FILING DATE: 10-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-4737

CC	ILLEKAK. (V13) JRU 0992.			
CC	INFORMATION FOR SEQ ID NO: 2:			
CC	SEQUENCE CHARACTERISTICS:			
CC	LENGTH: 2441 amino acids			
CC	TYPE: amino acid			
CC	TOPOLOGY: linear			
CC	MOLECULE TYPE: protein			
CC	SEQUENCE 2441 AA; 265473 MW; 30223014 CN;			
Query Match	2.1%	Score 205;	DB 2;	Length 2441;
Best Local Similarity	29.4%;	Pred. No. 1.34e-05;		
Matches	65;	Conservative	45;	Mismatches 97;
				Indels 14;
				Gaps 12;
Db	2116 MQPQGLQSOGMOPGPHHQPSL-QNUNMAQGVPRGPVPQPMAGLNPPQGALNI	2174		
Qy	1160 MRPTNTPKRLMQLRQLQQOFLNQSRQALELKMENTAGGAAYMRPMWQPQGFUNA	1219		
Db	2175 -MNPGNPNNTNN-PQYR-EMVRRQLQHQQQQQQQQQQQQQQNSASLAGMGHGSQF	2231		
Qy	1220 QMVARSRELLSHHFQRVRAMMMQSQQQQQQQQQQQQQQQQQQTQAFFSPPNVT-	1278		
Db	2232 QQPGGGGYA-PAMQQRMQHLPIQGS-SMGMAAP-MGQLG--Q-XGOPGLGADSTP	2284		
Qy	1279 ASFSMDGLLAGTPMPQAP-PQFPFYQNYGMGQPDPAFGRVSSPNNAMSSRMGFSQNP	1337		
Db	2285 NIQAALQORLIQQQQMKQ-QIGSPGQPNMSPQOHLMSGP	2324		
Qy	1338 MMQHPQAASIYSSEKMGWPSGNLARNSFSQQFAHQNP	1378		
RESULT	12			
ID	US-08-246-982A-6	STANDARD;	PRT;	3144 AA.
xx	xxxxxx			
AC				
xx				
DT				
XX	Sequence 6, Application US/08246982A			
DE	Sequence 6, Application US/08246982A			
CC				

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 19:18:26 ; Search time 1821.02 Seconds
(without alignments)
7852.012 Million cell updates/sec

Title: US-09-041-994-1
Perfect score: 4496
Sequence: 1 GCTGATGCTGGACTCAGAG.....CATTGAGCAGGAATCTAG 4496

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

abase :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*
16: gb_un.*
17: gb_vi.*
18: em_fun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pi.*
29: em_ro.*
30: em_sts.*
31: em_sy.*
32: em_un.*
33: em_vi.*
34: gb_htg1.*
35: gb_htg2.*
36: gb_in1.*
37: gb_in2.*
38: gb_ba1.*
39: gb_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4468.2	99.4	4495	11	AF010227 Homo sapi

ALIGNMENTS

RESULT 1

LOCUS	AF010227	4495 bp	mRNA	PRI	13-AUG-1997
DEFINITION	Homo sapiens receptor-associated coactivator 3 (RAC3) mRNA,				
ACCESSION	AF010227				
NID	g2318005				
VERSION	AF010227.1	GI:2318005			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
REFERENCE	1 (bases 1 to 4495)				
AUTHORS	Li, H., Gomes, P.J., and Chen, J.D.				
TITLE	RAC3, a steroid/nuclear receptor-associated coactivator that is related to SRC-1 and TIF2				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 94 (16), 8479-8484 (1997)				
MEDLINE	97385128				
REFERENCE	2 (bases 1 to 4495)				
AUTHORS	Chen, J.D., Li, H., and Gomes, P.J.				
TITLE	Direct Submission				
JOURNAL	Submitted (25-JUN-1997) Pharmacology and Molecular Toxicology, University of Massachusetts Medical School, 55 Lake Avenue North,				

```

FEATURES
  source          Worcester, MA 01655-0126, USA
                  Location/Qualifiers
                    1. .4495
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /cell_line="HeLa"
                    1. .4495
                      /gene="RAC3"
                      86. .4339
                      /gene="RAC3"
                      /note="transcriptional coactivator with intrinsic->histone
acetyltransferase activity; member of the steroid/nuclear
receptor-associated coactivator family which includes
steroid receptor coactivator 1 (SRC-1), transcriptional
intermediate factor 2 (TIF2), and receptor associated
coactivator 3; similar to the mouse p300/CBP/co-integrator
protein (P/CIP) and the activator of retinoid receptors
(ACR)"
                    /codon_start=1
                    /product="receptor-associated coactivator 3"
                    /protein_id="AAC51663.1"
                    /db_xref="PID:92318006"
                    /db_xref="GI:2318006"
                    /translation="MSGIGENLDP LASSRRKRLPCDTPGQGLTSGEKKRRREQESRY
IEELAEIISANLSDIDNFNVKDKCAILKETVROIROIKQGGKTSKDDVDQADVSS
TGQGVLDKSLGLDQLDGLFVNVREANIVFSENVQYLOQKDDLVNTSVYNI
LHEEDKDFLKNLPSKTVNGSVNTNETQOKSHTEFNCMLMKTPHDLIEDINASPENR
QRYETMOCFALSQPRAMEDGEDIQSCMICVARKITIGERTFFSPNPFSTIRHDLSEK
VNIIDTNSLRSPGMEEDIIIRCIQRFSLNDGQSWSKRHYQAYLNGHAETPVYR
FSLADGTIVTAQTSKLFKRPNTVNDHRGFVSTHFLQREONGYRPNPNVPGQIRPMA
GCNSVGGMSSNMQGLPSSRAYGLADPTTGOMSGARYGGSSNTASLTTPGPMOS
PSSYONNGLNLMSSPHGSPGLAPNOQNMIMSPRNRGSPKIASHOFSPVAGVHSPWA
SSGNTGNHSSSSLSLQAISEGVTGSLTSLSPGPKLDNPNMNIOTPSKVSNOQ
SKSPGTYCQNPVSESSQNSRDHLSDEKSSVEGAENQRGPLESKGRKLDQL
LTCSSDRGHSSSLNPLDSCKESSVSPTSQSVSSSTSGVSSVSSNMHGSLLQEKH
RILHLKLNSSPAEVAKITAETGKDTSSITSCGDGVYKQQLSPKKNENALLRY
LLDRDPSDALSKLPQVGVNDKMSQCTSSITPSSSQEKDPKIKYETSEEGSDJD
NLDAIIGLDTSSDFNLSISNGSLGTQKQVFGTNSLGLKSSQSVQSIIRPVNRAV
SLDSPSVSGSPPVKNTSAPFPLPKPQMLGNPRMDSQENYGSMSGPNRNTVTQT
PSSDWDGLPLSKAGRMENSMNSMGRFGDYNTSLRPLGGSITPLULKSNISLPGAR
PVLQOQOOLMQMRPEGLPMGMGANPYGOAAASNOLGSPDMLSMQVSHGTQNRPLL
RNSLDDLVGPPNSGDSERALLDLHTLSNTDATGLEIDRALGIPELYNQOAL
EPKQDAFGQEAAYVMQKAGLYGOTYPAQGPMPQGGFHLQGGSPSPSNMNMNOQG
VFPLQGMHPRANIRPTNTPKOLRMOLQORLOGOQFLNOSROALELKMENPRTAGGA
VMPMPQOQGFLENAQVORSRELLSHFERQEVANMQQOQQOQQOQQOQQOQQOQQ
OQOQQOQTAFSPPNPTVAFSDMCLLAGFTMQAPQQFPYQPNLNGQOQPPAFGR
VSPPNAMSSRMGFPQNPMMQHPQAASIIYQSEMKGWPNMPSGMGPDQKYC"
                  E COUNT      1354 a 1048 -c 1043 g 1050 t
ORIGIN

```

```

Query Match      99.48; Score 4468.2; DB 11; Length 4495;
Best Local Similarity 99.88; Pred. No. 0;
Matches 4486; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 GCTGGATGCTGATATATCAAGATGAGTGGATTAGAGAAACTTGGATCCACTGGC 120
DB 1 GCTGGATGCTGATATATCAAGATGAGTGGATTAGAGAAACTTGGATCCACTGGC 120
QY 61 AGCCAGTTGCTGATATATCAAGATGAGTGGATTAGAGAAACTTGGATCCACTGGC 120
DB 61 AGCCAGTTGCTGATATATCAAGATGAGTGGATTAGAGAAACTTGGATCCACTGGC 120
QY 121 CAGTATTCAGAAACGCAATGCTGATGATGATCTCCAGAGCAAGGCTTACCTGCG 180
DB 121 CAGTATTCAGAAACGCAATGCTGATGATGATGATCTCCAGAGCAAGGCTTACCTGCG 180
QY 181 TGGTGAACACGGACGGACGAGGAAGTAAATATATTGAAGATGGCTGACCTGAT 240
DB 181 TGGTGAACACGGACGGACGAGGAAGTAAATATATTGAAGATGGCTGACCTGAT 240
QY 241 ATCTGCCAATCTTAGTGATATTGACAATTTCAATGTCAACACGAGATAATGCGGATTT 300

```

```

|||||
241 ATCTGCCAATCTTAGTGATATTGACAATTTCAATGTCAACACGAGATAATGTCGATTTT 300
QY 301 AAAGGAACAGTAAAGACAGATACGTCATAAATAAAGAGCAAGGAAAACTATTTCCTAATGA 360
DB 301 AAAGGAACAGTAAAGACAGATACGTCATAAATAAAGAGCAAGGAAAACTATTTCCTAATGA 360
QY 361 TGATGATGTTTCAAAAGCCGATGATCTTCTACAGGCGAGGAGTATTATTGATAAGACTC 420
DB 361 TGATGATGTTTCAAAAGCCGATGATCTTCTACAGGCGAGGAGTATTATTGATAAGACTC 420
QY 421 CTTAGGACCGCTTTTACTTTCAGGCAATGGATGGTTTCTTATTTGGTGAATCGAGAGGC 480
DB 421 CTTAGGACCGCTTTTACTTTCAGGCAATGGATGGTTTCTTATTTGGTGAATCGAGAGGC 480
QY 481 AAACATTGTTATTGTTATCAGAAAATGTCACAAATACCTGCAATATAGCAAGAGGACCT 540
DB 481 AAACATTGTTATTGTTATCAGAAAATGTCACAAATACCTGCAATATAGCAAGAGGACCT 540
QY 541 GGTAAACACAGTCTTTTACAAATATCTTACATGAAGAGACAGAAAGGATTTCTTAAGAA 600
DB 541 GGTAAACACAGTCTTTTACAAATATCTTACATGAAGAGACAGAAAGGATTTCTTAAGAA 600
QY 601 TTTACCAAAATCTACAGTTAATGGAGTTTCTTGACAAATGAGCCCAAGACAAAAAAG 660
DB 601 TTTACCAAAATCTACAGTTAATGGAGTTTCTTGACAAATGAGACCCCAAGACAAAAAAG 660
QY 661 CCATACATTTAATTTGCCGTATGTTGATGAAAACACCAACATGATATTCTGGAAGACATAA 720
DB 661 CCATACATTTAATTTGCCGTATGTTGATGAAAACACCAACATGATATTCTGGAAGACATAA 720
QY 721 GGCAGTCTCTGAATGCGCAGAGATATGAACAATACAGTCTTTCCTGCTCTCAGCC 780
DB 721 GGCAGTCTCTGAATGCGCAGAGATATGAACAATACAGTCTTTCCTGCTCTCAGCC 780
QY 781 ACAGCTATGATGAGGAGGAGGAGGATTTGCAATCTTTGATGATCTGTGTGGCAGCGCG 840
DB 781 ACAGCTATGATGAGGAGGAGGAGGATTTGCAATCTTTGATGATCTGTGTGGCAGCGCG 840
QY 841 CATTTACTACAGGAGAAAGACATTTCCATCAAAACCTTGAGAGCTTTATTACAGACATGA 900
DB 841 CATTTACTACAGGAGAAAGACATTTCCATCAAAACCTTGAGAGCTTTATTACAGACATGA 900
QY 901 TCTTTTCAGGAAAGTCTGCAATATAGATACAAATTCAGTGAAGCTCTCCATGAGGCGCTGG 960
DB 901 TCTTTTCAGGAAAGTCTGCAATATAGATACAAATTCAGTGAAGCTCTCCATGAGGCGCTGG 960
QY 961 CTTTGAAGATATATCCGAAGGCTATTTCAGAGATTTTTTTAGTCTAAATGATGGCGAGTC 1020
DB 961 CTTTGAAGATATATCCGAAGGCTATTTCAGAGATTTTTTTAGTCTAAATGATGGCGAGTC 1020
QY 1021 ATGCTCCAGAAAGCTCACTATCAAGAGCTTATCTTAATGGCCATGCAGAAACCCGAGT 1080
DB 1021 ATGCTCCAGAAAGCTCACTATCAAGAGCTTATCTTAATGGCCATGCAGAAACCCGAGT 1080
QY 1081 ATATCGATTTCTGTTGGCTGATGGAATCTAGTGACTGCACAGACAAAAGCAAACTCTT 1140
DB 1081 ATATCGATTTCTGTTGGCTGATGGAATCTAGTGACTGCACAGACAAAAGCAAACTCTT 1140
QY 1141 CCGAAATCTCTGTAACAAATGATCGACATGGCTTTGTCTCAACCCACCTTCTTTCAGAGAGA 1200
DB 1141 CCGAAATCTCTGTAACAAATGATCGACATGGCTTTGTCTCAACCCACCTTCTTTCAGAGAGA 1200
QY 1201 ACAGATGAGATATAGACAAACCCAAATCTGTTGGACAAAGGATTTAGACCACCTATGGC 1260
DB 1201 ACAGATGAGATATAGACAAACCCAAATCTGTTGGACAAAGGATTTAGACCACCTATGGC 1260
QY 1261 TGGATGCAACAGTTCCGTAGCGGCATGAGTGTGTCGCCAAACCAAGCTTTACAGATGCC 1320
DB 1261 TGGATGCAACAGTTCCGTAGCGGCATGAGTGTGTCGCCAAACCAAGCTTTACAGATGCC 1320
QY 1321 GAGCAGCAGGCGCTTATGGCTTGGCAGACCTTAGCACCACAGGCGAGATGAGTGGAGGTAG 1380
DB 1321 GAGCAGCAGGCGCTTATGGCTTGGCAGACCTTAGCACCACAGGCGAGATGAGTGGAGGTAG 1380

```


Db 1321 GAGCAGAGGGCCCTATGGCTTGGCAGACCCTAGCACACAGGGCAGATGAGTGGAGCTAG 1380
Qy 1381 GTATGGGGTTCCAGTAACATAGCTTTCATGACCCCTGGGCCAGGCATGCAATCACCATC 1440
Db 1381 GTATGGGGTTCCAGTAACATAGCTTTCATGACCCCTGGGCCAGGCATGCAATCACCATC 1440
Qy 1441 TTCTTACCAGAACAACTATGGCTCAACATGAGTAGGCCCCCAATGGAGTCTCTGG 1500
Db 1441 TTCTTACCAGAACAACTATGGCTCAACATGAGTAGGCCCCCAATGGAGTCTCTGG 1500
Qy 1501 TCTTGGCCCAACAGCAGAAATATCATGATTTCTCTGTAATCGTGGAGTCCAAAGAT 1560
Db 1501 TCTTGGCCCAACAGCAGAAATATCATGATTTCTCTGTAATCGTGGAGTCCAAAGAT 1560
Qy 1561 AGCCTCAATCAGATTTCTCTGTTGCAAGTGTGCACCTCTCCATGGCATCTTCTGGCAA 1620
Db 1561 AGCCTCAATCAGATTTCTCTGTTGCAAGTGTGCACCTCTCCATGGCATCTTCTGGCAA 1620
Qy 1621 TACTGGGACACAGCTTTTCCAGAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
Db 1621 TACTGGGACACAGCTTTTCCAGAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
Qy 1681 TGTGGGACTTCCCTTTATCTCTCTCATCACCAGGCCCAAAATTTGGATACTCTCC 1740
Db 1681 TGTGGGACTTCCCTTTATCTCTCTCATCACCAGGCCCAAAATTTGGATACTCTCC 1740
Qy 1741 CAATATGAATATTACCCCAACCAAGTAAGTAAGCAATCAGGATTCCTCTGGG 1800
Db 1741 CAATATGAATATTACCCCAACCAAGTAAGTAAGCAATCAGGATTCCTCTGGG 1800
Qy 1801 CTTTATTTGACCAAAATTCAGTGGAGATTCATGTCAGTCAAAATACAGAGATCA 1860
Db 1801 CTTTATTTGACCAAAATTCAGTGGAGATTCATGTCAGTCAAAATACAGAGATCA 1860
Qy 1861 CCTCAGTGACAAAGAAAGTAAGGAGAGAGTGTGGGGGCGAGAGATCAAGGGGTCC 1920
Db 1861 CCTCAGTGACAAAGAAAGTAAGGAGAGAGTGTGGGGGCGAGAGATCAAGGGGTCC 1920
Qy 1921 TTTGAAAGCAAGGTCATAAAAAATTAAGTCAAGTACTTACCTGTTCTTCTGATGACCG 1980
Db 1921 TTTGAAAGCAAGGTCATAAAAAATTAAGTCAAGTACTTACCTGTTCTTCTGATGACCG 1980
Qy 1981 GGGTCATTCCTTTCAGCAACTCCCTAGATTCAGATTCAGATTCAGATTCAGATTCAG 2040
Db 1981 GGGTCATTCCTTTCAGCAACTCCCTAGATTCAGATTCAGATTCAGATTCAGATTCAG 2040
Qy 2041 TGTACCAAGCCCTCTGGAGTCTCTCTCTACATCTGGAGGATTCCTCTACATCCAA 2100
Db 2041 TGTACCAAGCCCTCTGGAGTCTCTCTCTACATCTGGAGGATTCCTCTACATCCAA 2100
Qy 2101 TATGATGGGTCACTGTTTACAGAGAACACCGGATTTTGCACAGATTCCTGCAAGATGG 2160
Db 2101 TATGATGGGTCACTGTTTACAGAGAACACCGGATTTTGCACAGATTCCTGCAAGATGG 2160
Qy 2161 GAATTCACCACTGAGGTAGCCAGATTTACGACAGCCACTGGGAAGACACAGCAG 2220
Db 2161 GAATTCACCACTGAGGTAGCCAGATTTACGACAGCCACTGGGAAGACACAGCAG 2220
Qy 2221 TATACCTCTTGTGGGACGGAATTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAGAA 2280
Db 2221 TATACCTCTTGTGGGACGGAATTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAGAA 2280
Qy 2281 GGAGAAATGCACTCTTATGATACCTCTGACAGGAGATTCCTAGTATGACACTCTC 2340
Db 2281 GGAGAAATGCACTCTTATGATACCTCTGACAGGAGATTCCTAGTATGACACTCTC 2340
Qy 2341 TAAAGAACTACAGCCCAAGTGGAGGAGTGGACAATAAATGATCAGTCCACAGCTC 2400
Db 2341 TAAAGAACTACAGCCCAAGTGGAGGAGTGGACAATAAATGATCAGTCCACAGCTC 2400
Qy 2401 CACCATCTTACCTAGCTCAAGTCAAGAGAAAGCCCTTAAATTAAGCAGAGACAGTGAAG 2460
Db 2401 CACCATCTTACCTAGCTCAAGTCAAGAGAAAGCCCTTAAATTAAGCAGAGACAGTGAAG 2460

Qy 2461 GGGATCTGGAGACTTGGATAATCTAGATGCTATCTTTGGTATCTGACTAGTCTGACTT 2520
Db 2461 GGGATCTGGAGACTTGGATAATCTAGATGCTATCTTTGGTATCTGACTAGTCTGACTT 2520
Qy 2521 TTACAATAATTCATATCTCTCAAAATGGTAGTCACTCTGGGACTAAGCAACAGGTGTTCA 2580
Db 2521 TTACAATAATTCATATCTCTCAAAATGGTAGTCACTCTGGGACTAAGCAACAGGTGTTCA 2580
Qy 2581 AGGAACATAATCTCTGGTGTGAAAGTTCACAGTCTGTGAGTCTATTCCTCTCCATA 2640
Db 2581 AGGAACATAATCTCTGGTGTGAAAGTTCACAGTCTGTGAGTCTATTCCTCTCCATA 2640
Qy 2641 TAACCGAGCAGTGTCTCTGGATAGCCCTGTTCTGTTGGCTCAAGTCTCTCAGTAAAAA 2700
Db 2641 TAACCGAGCAGTGTCTCTGGATAGCCCTGTTCTGTTGGCTCAAGTCTCTCAGTAAAAA 2700
Qy 2701 TATCAGTGTCTTCTCCCATGTTTACCAAGAACCCATGTTGGGTGGGAATCCAAAGATGAT 2760
Db 2701 TATCAGTGTCTTCTCCCATGTTTACCAAGAACCCATGTTGGGTGGGAATCCAAAGATGAT 2760
Qy 2761 GGATAGTCAGGAAATATGAGTCAAGTATGGGTGGCCAAACCGAAATGAGTGTGAC 2820
Db 2761 GGATAGTCAGGAAATATGAGTCAAGTATGGGTGGCCAAACCGAAATGAGTGTGAC 2820
Qy 2821 TCAGACTCTTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGAGAAATGGAAAC 2880
Db 2821 TCAGACTCTTCTCAGGAGACTGGGCTTACCAAACTCAAAAGCCGCGAGAAATGGAAAC 2880
Qy 2881 TATGAATTCAACTCCATGGGAAGACCCAGGAGAGATTAATACTTCTTTACCCAGACC 2940
Db 2881 TATGAATTCAACTCCATGGGAAGACCCAGGAGAGATTAATACTTCTTTACCCAGACC 2940
Qy 2941 TGCAGTGGGTGGCTCTATTCACATTCCTCTCTGCTTAAATAGCATACCAAGTGGAG 3000
Db 2941 TGCAGTGGGTGGCTCTATTCACATTCCTCTCTGCTTAAATAGCATACCAAGTGGAG 3000
Qy 3001 ACCAGTATTCACAGCAGCAGCAGATGCTTCAAAATGAGGCTTGGTAAATCCCATGGG 3060
Db 3001 ACCAGTATTCACAGCAGCAGCAGATGCTTCAAAATGAGGCTTGGTAAATCCCATGGG 3060
Qy 3061 AATGGGGCTTAATCCCTATGGCCAAAGCAGCATCTTAACCAACTGGGTTCTTGGCCGA 3120
Db 3061 AATGGGGCTTAATCCCTATGGCCAAAGCAGCATCTTAACCAACTGGGTTCTTGGCCGA 3120
Qy 3121 TGGCATGTGTGTCATGAACAAGTTCCTGACACACAAATAGGCTCTCTTCTTAGGAA 3180
Db 3121 TGGCATGTGTGTCATGAACAAGTTCCTGACACACAAATAGGCTCTCTTCTTAGGAA 3180
Qy 3181 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTACGAAAGAGC 3240
Db 3181 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGAAGCCAGAGTACGAAAGAGC 3240
Qy 3241 ATTATTTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGAGGCTTGGAGAAAT 3300
Db 3241 ATTATTTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGAGGCTTGGAGAAAT 3300
Qy 3301 TGACAGAGCTTGGGCAATTCCTGACATGTCATCAGGACAGGATTTAGAGCCCAACA 3360
Db 3301 TGACAGAGCTTGGGCAATTCCTGACATGTCATCAGGACAGGATTTAGAGCCCAACA 3360
Qy 3361 GGATGCTTTTCAAGGCCAAGAACGACAGATTAATGATGGATCAGAGGAGGATTTATGG 3420
Db 3361 GGATGCTTTTCAAGGCCAAGAACGACAGATTAATGATGGATCAGAGGAGGATTTATGG 3420
Qy 3421 ACAGACATACCCAGCAGAGGCGCTTCCAAATGCAAGGAGGCTTCTCATCTTCAGGACATC 3480
Db 3421 ACAGACATACCCAGCAGAGGCGCTTCCAAATGCAAGGAGGCTTCTCATCTTCAGGACATC 3480
Qy 3481 ACCATCTTTTAACTCTATGATGAATCAGATGAACAGCAGGCAATTTCTCTTCCAGG 3540
Db 3481 ACCATCTTTTAACTCTATGATGAATCAGATGAACAGCAGGCAATTTCTCTTCCAGG 3540

```
QY 3541 AATGACCCACGAGCCAAACATCATGAGACCCCGGACAAACACCCCAAGCAACTTAGAAT 3600
|||||
Db 3541 AATGACCCACGAGCCAAACATCATGAGACCCCGGACAAACACCCCAAGCAACTTAGAAT 3600
QY 3601 GCAGCTTCAGACAGAGGCTGAGGGCCAGCAGTGTGTTTGAATCAGAGCGGACAGCACTTGA 3660
|||||
Db 3601 GCAGCTTCAGACAGAGGCTGAGGGCCAGCAGTGTGTTTGAATCAGAGCGGACAGCACTTGA 3660
QY 3661 ATTGAATGGAACACCTACTGCTGTGTGCTGCGGTGATGAGGCTATGATGAGCC 3720
|||||
Db 3661 ATTGAATGGAACACCTACTGCTGTGTGCTGCGGTGATGAGGCTATGATGAGCC 3720
QY 3721 CCAGCAGGGTTCCTTAATGCTCAATGCTGCGCCACAGCAGCAGAGAGTGTCTAAGTCA 3780
|||||
Db 3721 CCAGCAGGGTTCCTTAATGCTCAATGCTGCGCCACAGCAGCAGAGAGTGTCTAAGTCA 3780
QY 3781 TCACATCCGACACAGAGGTGCTATGATGATGACAGCAGCAACAG-----CAGCA 3834
|||||
Db 3781 TCACATCCGACACAGAGGTGCTATGATGATGACAGCAGCAACAGCAACAGCA 3840
3835 GCAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCAACAGCAACAGCCAGCC 3894
|||||
Db 3841 GCAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCAACAGCAACAGCCAGCC 3900
QY 3895 CTTACAGCCACCTCCTTAATGCTGCTTCCCGCAGCATGGATGGGCTTTTGGCAGGACC 3954
|||||
Db 3901 CTTACAGCCACCTCCTTAATGCTGCTTCCCGCAGCATGGATGGGCTTTTGGCAGGACC 3960
QY 3955 CACAATGCCACAGCTCCTCCGCAACAGTTTCCATATCAACCAAAATATGGAATGGACA 4014
|||||
Db 3961 CACAATGCCACAGCTCCTCCGCAACAGTTTCCATATCAACCAAAATATGGAATGGACA 4020
QY 4015 ACACACAGATCCAGCCTTTGGTGCAGTGTCTAGTCTCTCCCAATGCAATGATGCTGCAAG 4074
|||||
Db 4021 ACACACAGATCCAGCCTTTGGTGCAGTGTCTAGTCTCTCCCAATGCAATGATGCTGCAAG 4080
QY 4075 AATGGGTCCCTCCAGAAATCCCATGATGCAACACCCCGAGGCTGCATCCATCTATCAGTC 4134
|||||
Db 4081 AATGGGTCCCTCCAGAAATCCCATGATGCAACACCCCGAGGCTGCATCCATCTATCAGTC 4140
QY 4135 CTCAGAAATGAAGGCTGCGCATCAGGAAATTTGGCCAGGAAACAGCTCTTTTCCAGCA 4194
|||||
Db 4141 CTCAGAAATGAAGGCTGCGCATCAGGAAATTTGGCCAGGAAACAGCTCTTTTCCAGCA 4200
QY 4195 GCAGTTTGGCCACAGGGGAATCCTGCAGTGTATAGTATGTCACATGATGCGCAGCAG 4254
|||||
Db 4201 GCAGTTTGGCCACAGGGGAATCCTGCAGTGTATAGTATGTCACATGATGCGCAGCAG 4260
4255 TGGTACATGGGACAGATGAACATGAACCCCATGCCCATGCTGCGCATGCTATGGGTCC 4314
|||||
Db 4261 TGGTACATGGGACAGATGAACATGAACCCCATGCCCATGCTGCGCATGCTATGGGTCC 4320
QY 4315 TGATCAGAAATACCTGTCAGATCTCTGCACAGACCTCTTAAGGAAACACACTGTACAAA 4374
|||||
Db 4321 TGATCAGAAATACCTGTCAGATCTCTGCACAGACCTCTTAAGGAAACACACTGTACAAA 4380
QY 4375 TGACACTGCACTAGGATATTGGGAAGGAATCATTTGTCAGGCACTTCCTTGGAGAA 4434
|||||
Db 4381 TGACACTGCACTAGGATATTGGGAAGGAATCATTTGTCAGGCACTTCCTTGGAGAA 4440
QY 4435 AGGACACGCTTGGCTCCATCAAGGGTATTTTAAGTGTGTCATTTGAGCAGGA 4489
|||||
Db 4441 AGGACACGCTTGGCTCCATCAAGGGTATTTTAAGTGTGTCATTTGAGCAGGA 4495

RESULT 2
AF012108
LOCUS AF012108 6835 bp mRNA PRI 21-AUG-1997
DEFINITION Homo sapiens Amplified in Breast Cancer (AIB1) mRNA, complete cds.
ACCESSION AF012108
NID G2331249
VERSION AF012108.1 GI:2331249
KEYWORDS,
```

```
SOURCE
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS      Anzick,S.L., Kononen,J., Walker,R.L., Azorsa,D.O., Tanner,M.M.,
              Guan,X.Y., Sauter,G., Kallioniemi,O.P., Trent,J.M. and Meltzer,P.S.
TITLE        AIB1, a steroid receptor coactivator amplified in breast and
              ovarian cancer
JOURNAL       Science 277 (5328), 965-968 (1997)
MEDLINE       97400625
REFERENCE     2 (bases 1 to 6835)
AUTHORS      Anzick,S.L., Walker,R.L., Guan,X.-Y. and Meltzer,P.S.
TITLE        Direct Submission
JOURNAL       Submitted (01-JUL-1997) Laboratory of Cancer Genetics, National
              Human Genome Research Institute, 49 Convent Dr., Bethesda, MD
              20892-4470, USA
FEATURES
             Location/Qualifiers
             1..6835
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="20"
                /map="20q12"
             1..6835
                /gene="AIB1"
                201..4463
                /gene="AIB1"
                /note="nuclear receptor co-activator"
                /codon_start=1
                /product="Amplified in Breast Cancer"
                /protein_id="AAC51677.1"
                /db_xref="PID:g2331250"
                /db_xref="GI:2331250"
                /translation="MSGLEGENLDPLASDRKKLPDTPGQGLTCSGKRRRQESKY
              IBEAELISANLSDIDNFVDPKAILKTRVQIKKEQKTIISNDDVDQKADVSS
              TQGVIDKDSLGLLOALDGLFVYNRDGNIVFVSENYTOYLOKLEDLVNTSVNI
              LHEEDKDFLKNLPKSTVNGSVNTETORQKSHFTNCRLMKTPHDIEDINASPEMR
              QREYTMQCFALSPRAMMEGEDLOSCMTCVARRITTGERTFSPESPETITRDLSGK
              VYNIITNSLRSMRPFEDIIRRCIQREFSLNDGGSQKRYQAYLNHGAETPVYR
              FSLADGTIVTAQTKSLFRNPVINDRHGTVSTHFLQREQNGYRPNPVGGQIRPMPA
              GCNSVGGMSMSPNQGLQMPSSRAYGLADPTTQMGSGARYGGSNIGIASLTGPGMQS
              PSSYQNNYGLNMSPPHSGPLAPNQNMISIPRNGSPKIASHQSPVAVHSPMA
              SSNGTGNHFSFSSLSALQATSEGVGTSLTSLSSPGKLDNSPNNITQPSKVSNOD
              SKSPGLFYCDQNPVSSMCSNRDLSLDSCKESSVYTSFGSVSSTSNMGSLLOEKH
              LTCSDDRGHSLTNSPLDSCCKESSVYTSFGSVSSTSNMGSLLOEKH
              RLHLKLNQNSPAEVAKITAEATKDTSSITSCGDGNVYKQESLPKKNALLRY
              LLDKDDPSDALSKELQPVQGVNKMSTQSTTIPSSOEKDPKIKTETSEBGGDL
              NLDTALIGDLTSSDFYNNISSTSSNGSHLGTQQVFGQTNLSGLKSSVSQSVIRPPYNRV
              LUDSPVSVSSSPVKNISAFPMKPMLGPNRMDQENYSSMGSGPNRNVITOT
              PSSGDWGLPNSKAGRMENSMRPGGDYNTSLPRPALGGS IPTLPVLSNSIPGAR
              PVLQOQOQLQMRPGEIPMGGANPYGQAAASNQLGSHWPDGMSLSEVQSHGTQNRPLL
              RNSLDLVGPPSNLEGGQSDERALLQLHLTLNLTNDATGLEEDRALGIPELVYNGOAL
              EPKDAFQGGQEAAMVMDQAGLYGQTYPAQGPMPQGGHLOQGSFNSMNMNOQG
              NPLQAGMHPRANIMRPTNPKQLRMQLQRLQGGQFLNQSRQALELKMENPTAGGAA
              VNRPMQOQGFNAQMAQRSLRLLSHHFROORVAMMMQOQOQOQOQOQOQOQOQOQ
              OQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ
              OQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ
              FGRVSSPPNAMSSRMSPQNPMQHPQAAISYQSSEKMGWPSGNLARNSSFSQQQFA
              HOGNPAVYSVHMNNGSSGHGMQNMMPMSGMPMGPDQKIC"
BASE COUNT   1988 a 1479 c 1528 g 1840 t
ORIGIN
```

```
Query Match      99.0%; Score 4451.6; DB 11; Length 6835;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 4482; Conservative 0; Mismatches 9; Indels 15; Gaps 1;

QY 1 GCTGGATGTTGGACTCAGACCAATAAAAAATAACTGCTTGACATCCTTTGACGTGTT 60
|||||
Db 116 GCTGGATGTTGGACTCAGACCAATAAAAAATAACTGCTTGACATCCTTTGACGTGTT 175
|||||
QY 61 AGCCAGTTGCTGATGATATTCAGATGAGTGGATTAGGAGAAAACCTTGGATCCACTGGC 120
|||||
Db 176 AGCCAGTTGCTGATGATATTCAGATGAGTGGATTAGGAGAAAACCTTGGATCCACTGGC 235
```

QY 121 CAGTGATTCAGAAAACCCAAATTCGCCATGTGATCTCCAGGACAAGGCTCTTACTGCGAG 180
DB 236 CAGTGATTCAGAAAACCCAAATTCGCCATGTGATCTCCAGGACAAGGCTCTTACTGCGAG 295
QY 181 TGGTGA AAAACCGGAGACGGGAGCAGGAAGTAATATATATGAAGAAATTTGGCTGAGCTGAT 240
DB 296 TGGTGA AAAACCGGAGACGGGAGCAGGAAGTAATATATATGAAGAAATTTGGCTGAGCTGAT 355
QY 241 ATCTGCCAATCTTAGTGATATTTGCAATTTTCAATGTCTCAACCCAGATAAATTTGCGATTTT 300
DB 356 ATCTGCCAATCTTAGTGATATTTGCAATTTTCAATGTCTCAACCCAGATAAATTTGCGATTTT 415
QY 301 AAAGGAACAGTAAGACAGATACGTCAATTAAGAGAGCAAGGAAGAACTATTTCCAAATGA 360
DB 416 AAAGGAACAGTAAGACAGATACGTCAATTAAGAGAGCAAGGAAGAACTATTTCCAAATGA 475
QY 361 TGATGATGTTCAAAAAGCCGATGTATCTCTACAGGCGAGGAGTTATTGATAAGACTC 420
DB 476 TGATGATGTTCAAAAAGCCGATGTATCTCTACAGGCGAGGAGTTATTGATAAGACTC 535
QY 421 CTTAGGACCGCTTTTACTTCAGGCATTTGGATGGTTTCTTATTTGTGTGAATCGAGAGGC 480
DB 536 CTTAGGACCGCTTTTACTTCAGGCATTTGGATGGTTTCTTATTTGTGTGAATCGAGAGGC 595
QY 481 AAACATTTGATTTGATCAGAAAATGTCCACAAATACCTGCAATATATAGCAAGAGACCT 540
DB 596 AAACATTTGATTTGATCAGAAAATGTCCACAAATACCTGCAATATATAGCAAGAGACCT 655
QY 541 GGTTAACACAGTGTTTACATATCTTACATGAAGAGACAGAAAGGATTTTCTTAAGAA 600
DB 656 GGTTAACACAGTGTTTACATATCTTACATGAAGAGACAGAAAGGATTTTCTTAAGAA 715
QY 601 TTTACCAAAATCTACAGTTAATGGAGTTTCTCGCAAAATGAGCCCAAGCAACAAAAAG 660
DB 716 TTTACCAAAATCTACAGTTAATGGAGTTTCTCGCAAAATGAGCCCAAGCAACAAAAAG 775
QY 661 CCATACATTTAATTTGCCGTATGTTGATGA AAAACACCACTGATATTTCTGGAAGACATAAA 720
DB 776 CCATACATTTAATTTGCCGTATGTTGATGA AAAACACCACTGATATTTCTGGAAGACATAAA 835
QY 721 CGCCAGTCTGAATGCGCCAGAGATATGAACAAATGCAAGTTCCTTGGCCCTGTCTCAGCC 780
DB 836 CGCCAGTCTGAATGCGCCAGAGATATGAACAAATGCAAGTTCCTTGGCCCTGTCTCAGCC 895
QY 781 ACAGAGCTATGATGAGGAAGGGAAGATTTGCAATCTTGTATGATCTGTGTGGCAGCGCG 840
DB 896 ACAGAGCTATGATGAGGAAGGGAAGATTTGCAATCTTGTATGATCTGTGTGGCAGCGCG 955
QY 841 CATTACTACAGGAAGAAGACATTTCCATCAAAACCCCTGAGAGCTTTATTACCAGACATGA 900
DB 956 CATTACTACAGGAAGAAGACATTTCCATCAAAACCCCTGAGAGCTTTATTACCAGACATGA 1015
QY 901 TCTTTACAGGAAGGTTGTCATATATAGATACAAATTCAGTACAGTCCCTCAGAGCCCTGG 960
DB 1016 TCTTTACAGGAAGGTTGTCATATATAGATACAAATTCAGTACAGTCCCTCAGAGCCCTGG 1075
QY 961 CTTTGAAGATATAATCCGAAGGTTATTTCAGAGATTTTATTAGTCTTAATGATGGCAGCTC 1020
DB 1076 CTTTGAAGATATAATCCGAAGGTTATTTCAGAGATTTTATTAGTCTTAATGATGGCAGCTC 1135
QY 1021 ATGTTCCAGAAACGCTACTATCAAGAGCTTATCTTAATGGCCATGCAGAAAACCCAGT 1080
DB 1136 ATGTTCCAGAAACGCTACTATCAAGAGCTTATCTTAATGGCCATGCAGAAAACCCAGT 1195
QY 1081 ATATCGATTTCTGTTGGCTGATGGAACATATAGTACTGCACAGACAAAACCAACTCTT 1140
DB 1196 ATATCGATTTCTGTTGGCTGATGGAACATATAGTACTGCACAGACAAAACCAACTCTT 1255
QY 1141 CCGAAATCTCTAACAATATGACATGGCTTTGCTCTCAACCCACTTCCCTTCAGAGAGA 1200
DB 1256 CCGAAATCTCTAACAATATGACATGGCTTTGCTCTCAACCCACTTCCCTTCAGAGAGA 1315

QY 1201 ACAGAAATGGATATATAGACCAAAACCCAAATTCCTGTTGGACAAGGGATTTAGACCACCTATGGC 1260
DB 1316 ACAGAAATGGATATATAGACCAAAACCCAAATTCCTGTTGGACAAGGGATTTAGACCACCTATGGC 1375
QY 1361 TGGATGAACAAGTTTCGGTAGGCGCATGAGTATCTGCGCAAAACCAAGGCTTTACAGATGCC 1320
DB 1376 TGGATGAACAAGTTTCGGTAGGCGCATGAGTATCTGCGCAAAACCAAGGCTTTACAGATGCC 1435
QY 1321 GAGCAGCAGGSCCTTATGGCTTTGGCAGACCCCTAGCACCACAGGGCAGATGAGTGAGGCTAG 1380
DB 1436 GAGCAGCAGGSCCTTATGGCTTTGGCAGACCCCTAGCACCACAGGGCAGATGAGTGAGGCTAG 1495
QY 1381 GTATGGGGTTCCAGTAACATAGCTTCATGACCCCTGGGCGCAGCATGCAATCACCATC 1440
DB 1496 GTATGGGGTTCCAGTAACATAGCTTCATGACCCCTGGGCGCAGCATGCAATCACCATC 1555
QY 1441 TTCTTACAGAAACAACATATGGCTCAACATGAGTAGCCCCCACATGGAGTGCCTGG 1500
DB 1556 TTCTTACAGAAACAACATATGGCTCAACATGAGTAGCCCCCACATGGAGTGCCTGG 1615
QY 1501 TCTTGGCCCCAACAGCAGATAATCATGATTTCTCTCGTAATCGTGGAGTCCAAAGAT 1560
DB 1616 TCTTGGCCCCAACAGCAGATAATCATGATTTCTCTCGTAATCGTGGAGTCCAAAGAT 1675
QY 1561 AGCTCATACATCAGTTTCTCTGTGCAAGTGTGCATCTCTCCCATGCGCATCTTCTGGCAA 1620
DB 1676 AGCTCATACATCAGTTTCTCTGTGCAAGTGTGCATCTCTCCCATGCGCATCTTCTGGCAA 1735
QY 1621 TACTGGGAACCCAGCTTTTCCAGCAGCTCTCTCAGTGCCCTGCAAGCCATCAGTGAAG 1680
DB 1736 TACTGGGAACCCAGCTTTTCCAGCAGCTCTCTCAGTGCCCTGCAAGCCATCAGTGAAG 1795
QY 1681 TGTGGGACTTCCCTTTTATCTACTGTCTATCACAGGCCCCAAATTTGGATTAATCTCTCC 1740
DB 1796 TGTGGGACTTCCCTTTTATCTACTGTCTATCACAGGCCCCAAATTTGGATTAATCTCTCC 1855
QY 1741 CAATATGAATATATACCAACCAAGTAAGTAAGCAATCAGGATTTCCAAGAGTCTCTCTGG 1800
DB 1856 CAATATGAATATATACCAACCAAGTAAGTAAGCAATCAGGATTTCCAAGAGTCTCTCTGG 1915
QY 1801 CTTTATTTGGCCAAACCAATCCAGTGGAGAGTTCAATGTGTGTCAGTCAATATACAGAGATCA 1860
DB 1916 CTTTATTTGGCCAAACCAATCCAGTGGAGAGTTCAATGTGTGTCAGTCAATATACAGAGATCA 1975
QY 1861 CCTCAGTGACAAAAGAAAGTAAGAGAGCAGTGTGAGGGGCGAGAGATCAAAAGGGTCC 1920
DB 1976 CCTCAGTGACAAAAGAAAGTAAGAGAGCAGTGTGAGGGGCGAGAGATCAAAAGGGTCC 2035
QY 1921 TTTGGAAGCAAAAGGTCATAAAAAATTTACTGCAAGTACTTACCTGTCTTCTGTGATGACC 1980
DB 2036 TTTGGAAGCAAAAGGTCATAAAAAATTTACTGCAAGTACTTACCTGTCTTCTGTGATGACC 2095
QY 1981 GGGTCATCTCTCTTGACCAACTCCCCCTAGATTCAGTTGTAAGAAATCTTCTGTAG 2040
DB 2096 GGGTCATCTCTCTTGACCAACTCCCCCTAGATTCAGTTGTAAGAAATCTTCTGTAG 2155
QY 2041 TGTCAACAGCCCTCTGAGTCTCTCTPACATCTGAGAGATATCTCTACATCAAA 2100
DB 2156 TGTCAACAGCCCTCTGAGTCTCTCTPACATCTGAGAGATATCTCTACATCAAA 2215
QY 2101 TATGATGGGTCACTGTTTACAAGAAAGACCCGGATTTTGCAAGTGTGTCAGAAATGG 2160
DB 2216 TATGATGGGTCACTGTTTACAAGAAAGACCCGGATTTTGCAAGTGTGTCAGAAATGG 2275
QY 2161 GAATTCACAGCTGAGGTAGCCAAAGATTTACTGCAACAGCCACTGGGAAGACACACGAG 2220
DB 2276 GAATTCACAGCTGAGGTAGCCAAAGATTTACTGCAAGACCCACTGGGAAGACACACGAG 2335
QY 2221 TATAACTTCTTGTGGGACCGAAATGTTGTAACAGCAGGAGCTTAAGTCTCTAAGAA 2280
DB 2336 TATAACTTCTTGTGGGACCGAAATGTTGTAACAGCAGGAGCTTAAGTCTCTAAGAA 2395
QY 2281 GGAGAAATAATGCATTTCTTAGATACCTGCTGGACAGGGATGATCCTAGTATGATGATCTC 2340

|||||
Db 2396 GGAGAAATATGCACTCTTATAGATACCTGCTGGACAGGGATGATCCTAGTGATGCACTCTC 2455
|||||
QY 2341 TAAAGAACTACAGCCCAAGTGAAGAGTGGACATAAATAAGTCAAGTGCACAGCTC 2400
|||||
Db 2456 TAAAGAACTACAGCCCAAGTGAAGAGTGGATATAAATAAGTCAAGTGCACAGCTC 2515
|||||
QY 2401 CACCATTCTTCAAGTCAAGTCAAGAGAAAGACCTTAAATTAAGACAGAGACAGTGAAGA 2460
|||||
Db 2516 CACCATTCTTCAAGTCAAGTCAAGAGAAAGACCTTAAATTAAGACAGAGACAGTGAAGA 2575
|||||
QY 2461 GGGATCTGGAGACTTGGATTAATCTAGATGCTATCTTGGTGATCTGACTAGTCTGACTT 2520
|||||
Db 2576 GGGATCTGGAGACTTGGATTAATCTAGATGCTATCTTGGTGATCTGACTAGTCTGACTT 2635
|||||
QY 2521 TTACAATAATTCATATCTTCAATGGTAGTCACTGTGAGCTATCTGCTCCATTA 2580
|||||
Db 2636 TTACAATAATTCATATCTTCAATGGTAGTCACTGTGAGCTATCTGAGCTTGAACAACAGTGTTC 2695
|||||
QY 2581 AGGAACCTAATCTCTGGGTTTGAAGTTCACAGTCTGTGAGCTATCTGCTCCATTA 2640
|||||
Db 2696 AGGAACCTAATCTCTGGGTTTGAAGTTCACAGTCTGTGAGCTATCTGCTCCATTA 2755
|||||
QY 2641 TAACCGAGAGCTGCTCTGGATAGCCCTGTTCTTGGTGGCTCAAGTCTCCAGTAAAAA 2700
|||||
Db 2756 TAACCGAGAGCTGCTCTGGATAGCCCTGTTCTTGGTGGCTCAAGTCTCCAGTAAAAA 2815
|||||
QY 2701 TATCAGTCTTCCCTCCATGTTTACAAAGCAACCCATGTTGGTGGGAATCCAAAGATGAT 2760
|||||
Db 2816 TATCAGTCTTCCCTCCATGTTTACAAAGCAACCCATGTTGGTGGGAATCCAAAGATGAT 2875
|||||
QY 2761 GGATAGTCAGGAAATATTCGCTCAAGTATGGTGGGCCAAACCGAAATGTGACTGTGAC 2820
|||||
Db 2876 GGATAGTCAGGAAATATTCGCTCAAGTATGGTGGGCCAAACCGAAATGTGACTGTGAC 2935
|||||
QY 2821 TCAGACTCTCTCCAGGAGACTGGGCTTACCAAACTCAAGGCCGGGAGAGTGAAC 2880
|||||
Db 2936 TCAGACTCTCTCCAGGAGACTGGGCTTACCAAACTCAAGGCCGGGAGAGTGAAC 2995
|||||
QY 2881 TATGAATTCAACTCCATGGAAGACAGGAGAGATTAATTAATCTTACCCAGAC 2940
|||||
Db 2996 TATGAATTCAACTCCATGGAAGACAGGAGAGATTAATTAATCTTACCCAGAC 3055
|||||
QY 2941 TGCACTGGGTGGCTTATTCCTCCATGCTCTTCGCTCTTAATAGATACCAAGTGGAG 3000
|||||
Db 3056 TGCACTGGGTGGCTTATTCCTCCATGCTCTTCGCTCTTAATAGATACCAAGTGGAG 3115
|||||
QY 3001 ACCAGTATTCGAACAGCAGCAGATGCTTCAATAGAGCCCTGGTGAATCCCATGGG 3060
|||||
Db 3116 ACCAGTATTCGAACAGCAGCAGATGCTTCAATAGAGCCCTGGTGAATCCCATGGG 3175
|||||
QY 3061 ATGGGGGCTAATCCCTATGGCCAAAGCAGCAGATCTAACCACTGGGTCTGGCCGA 3120
|||||
Db 3176 ATGGGGGCTAATCCCTATGGCCAAAGCAGCAGATCTAACCACTGGGTCTGGCCGA 3235
|||||
QY 3121 TGGCATCTTCTGCATGAACAAGTTCTCATGGCACTCAAAATAGGCTCTCTTAGGAA 3180
|||||
Db 3236 TGGCATCTTCTGCATGAACAAGTTCTCATGGCACTCAAAATAGGCTCTCTTAGGAA 3295
|||||
QY 3181 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGGAAGGCCAGAGTGAACAAAGAGC 3240
|||||
Db 3296 TTCCCTGGATGATCTTGTGGGCCACCTTCCAACTGGGAAGGCCAGAGTGAACAAAGAGC 3355
|||||
QY 3241 ATTATGGGACAGTGGACACTCTTCTCAGCAACACAGATGGGACAGGCTTGAAGAAAT 3300
|||||
Db 3356 ATTATGGGACAGTGGACACTCTTCTCAGCAACACAGATGGGACAGGCTTGAAGAAAT 3415
|||||
QY 3301 TGACAGAGCTTGGGCAATCTTGAACCTTCAATCAGGGACAGCATTAGAGCCCAACA 3360
|||||
Db 3416 TGACAGAGCTTGGGCAATCTTGAACCTTCAATCAGGGACAGCATTAGAGCCCAACA 3475
|||||
QY 3361 GGATGCTTCCAAAGGCCAAGACAGCAGTAAATGATGGATCAGAAGCAGGATATATGG 3420
|||||

Db 3476 GGATGCTTTCCAAAGGCCAAAGACAGCAGTAATATGATGATCAGAAGCAGGATATATGG 3535
QY 3421 ACAGACATACCCAGACAGGGCCCTCAATGCAAGAGAGCTTTCATCTTCAGGCAATC 3480
|||||
Db 3536 ACAGACATACCCAGACAGGGCCCTCAATGCAAGAGAGCTTTCATCTTCAGGCAATC 3595
|||||
QY 3481 ACCATCTTTAACTCTATGATGAATGAAACAGACAGGCAATTTTCCTCTCCAAG 3540
|||||
Db 3596 ACCATCTTTAACTCTATGATGAATGAAACAGACAGGCAATTTTCCTCTCCAAG 3655
|||||
QY 3541 ATGCAACCCAGCAGCAACATCATGAGACCCCGGACAAACACCCCAAGCAACTTGAAT 3600
|||||
Db 3656 AATGCAACCCAGCAGCAACATCATGAGACCCCGGACAAACACCCCAAGCAACTTGAAT 3715
|||||
QY 3601 GCAGCTTTCAGCAGAGGCTCAGGGCCAGCAGTTTGAATCAGAGCCGACAGCACTTGA 3660
|||||
Db 3716 GCAGCTTTCAGCAGAGGCTCAGGGCCAGCAGTTTGAATCAGAGCCGACAGCACTTGA 3775
|||||
QY 3661 ATTGAAATGGAACCCCTACTGCTGGTGTGCTGGGTGATGAGGCTTATGATGAGCC 3720
|||||
Db 3776 ATTGAAATGGAACCCCTACTGCTGGTGTGCTGGGTGATGAGGCTTATGATGAGCC 3835
|||||
QY 3721 CCAGCAGGCTTTTCTTAAATGCTCAAAATGCTGCCCAACCGACAGAGAGCTGCTAAGTCA 3780
|||||
Db 3836 CCAGCAGGCTTTTCTTAAATGCTCAAAATGCTGCCCAACCGACAGAGAGCTGCTAAGTCA 3895
|||||
QY 3781 TCACCTCCACACAGAGGCTGCTATGATGATGAGCAG-----CAGCA 3825
|||||
Db 3896 TCACCTCCACACAGAGGCTGCTATGATGATGAGCAGCAGCAGCAGCAACAGCA 3955
|||||
QY 3826 ACAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAACAGCAGCA 3885
|||||
Db 3956 GCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAGCAGCA 4015
|||||
QY 3886 AACCCAGGCTTTAGCCCACTCTTAATGCTGCTTCCCTCCAGCATGGATGGGTTTT 3945
|||||
Db 4016 AACCCAGGCTTTAGCCCACTCTTAATGCTGCTTCCCTCCAGCATGGATGGGTTTT 4075
|||||
QY 3946 GGCAGACCCCAACAGCAGCAGCAGCAGCAGCAGCAGTTTCCATATCAACCAATATGG 4005
|||||
Db 4076 GGCAGACCCCAACAGCAGCAGCAGCAGCAGCAGCAGTTTCCATATCAACCAATATGG 4135
|||||
QY 4006 AATGGCACAACACAGATCCAGCCTTTGCTGAGTGTCTAGTCTCCCAATCAATGAT 4065
|||||
Db 4136 AATGGCACAACACAGATCCAGCCTTTGCTGAGTGTCTAGTCTCCCAATCAATGAT 4195
|||||
QY 4066 GTGCTCAAGAATGGGTCCCTCCAGAAATCCCATGATGCAACCCCGAGCTGCATCCAT 4125
|||||
Db 4196 GTGCTCAAGAATGGGTCCCTCCAGAAATCCCATGATGCAACCCCGAGCTGCATCCAT 4255
|||||
QY 4126 CTATCAGTCTCAGAAATGAAGGCTGGCCATCAGGAAATTTGGCCAGGACAGCTCCTT 4185
|||||
Db 4256 CTATCAGTCTCAGAAATGAAGGCTGGCCATCAGGAAATTTGGCCAGGACAGCTCCTT 4315
|||||
QY 4186 TTCCCAAGCAGCAGTTTGGCCACCCAGGGGAATCCTGCAAGTGTATGATGTCACATGAA 4245
|||||
Db 4316 TTCCCAAGCAGCAGTTTGGCCACCCAGGGGAATCCTGCAAGTGTATGATGTCACATGAA 4375
|||||
QY 4246 TGGCAGCAGTGTACATGGGACAGATGAACATGAACCCCATGCCCCTGCTGGCATGCC 4305
|||||
Db 4376 TGGCAGCAGTGTGTACATGGGACAGATGAACATGAACCCCATGCCCCTGCTGGCATGCC 4435
|||||
QY 4306 TATGGTCTCTGATCAGAAATAGTCTGACATCTCTGCACAGGACCTCTTAAGGAAACCA 4365
|||||
Db 4436 TATGGTCTCTGATCAGAAATAGTCTGACATCTCTGCACAGGACCTCTTAAGGAAACCA 4495
|||||
QY 4366 CTCTAAGATGACACTGCACTAGGATTTATGGGAAGGAATCATTTGTCAGGACATCCATC 4425
|||||
Db 4496 CTGTACAAATGACACTGCACTAGGATTTATGGGAAGGAATCATTTGTCAGGACATCCATC 4555
|||||
QY 4426 TTGGGAAGAAAGGACAGCTTTGAGCTCCCATCAAGGGTATTTTAAAGTGTCTATTGAGC 4485
|||||
Db 4556 TTGGGAAGAAAGGACAGCTTTGAGCTCCCATCAAGGGTATTTTAAAGTGTCTATTGAGC 4615
|||||

QY 1021 ATGGTCCCAAGAACGTCACATCAAGAGCTTATCTTAATGGCCATGAGAAACCCAGT 1080
DB 1127 ATGGTCCCAAGAACGTCACATCAAGAGCTTATCTTAATGGCCATGAGAAACCCAGT 1186
QY 1081 ATATCGATTCTCGTTGGCTGATGGAACCTATAGTACTGCACAGACAAAAGCAAACTCTT 1140
DB 1187 ATATCGATTCTCGTTGGCTGATGGAACCTATAGTACTGCACAGACAAAAGCAAACTCTT 1246
QY 1141 CGGAATCTGTAAACAAATGATCGACATGGCTTTGTCTCAACCCACTTCCCTTCAGAGAGA 1200
DB 1247 CGGAATCTGTAAACAAATGATCGACATGGCTTTGTCTCAACCCACTTCCCTTCAGAGAGA 1306
QY 1201 ACAGAATGGATATAGACAAACCCAAATCTCTTGGACAAGGATATAGACACCTATATGGC 1260
DB 1307 ACAGAATGGATATAGACAAACCCAAATCTCTTGGACAAGGATATAGACACCTATATGGC 1366
QY 1261 TGGATGCAACAGTTCCGTAGGGGGCATGAGTATGTGCGCAACCAAGGCTTACAGATGCC 1320
DB 1367 TGGATGCAACAGTTCCGTAGGGGGCATGAGTATGTGCGCAACCAAGGCTTACAGATGCC 1426
QY 1321 GAGCAGCGGGCTATGGCTTGGCAGACCTAGCACCACAGGCGAGATGAGTGGAGCTAG 1380
DB 1427 GAGCAGCGGGCTATGGCTTGGCAGACCTAGCACCACAGGCGAGATGAGTGGAGCTAG 1486
QY 1381 GTATGGGGTCCAGTAACATAGCTTTCATTGACCCCTGGCCAGGCATGCAATCACCATC 1440
DB 1487 GTATGGGGTCCAGTAACATAGCTTTCATTGACCCCTGGCCAGGCATGCAATCACCATC 1546
QY 1441 TTCCATCCAGAACAACTATGGGCTCAACATGAGTAGCCGCCCAACCAAGGCTTACAGTCC 1500
DB 1547 TTCCATCCAGAACAACTATGGGCTCAACATGAGTAGCCGCCCAACCAAGGCTTACAGTCC 1606
QY 1501 TCTTGGCCCAACACAGCAATATCATGTTCTCTCGTATCTGGAGTCCAAAGAT 1560
DB 1607 TCTTGGCCCAACACAGCAATATCATGTTCTCTCGTATCTGGAGTCCAAAGAT 1666
QY 1561 AGCCTCACATCAGTTTCTCTGTTGCAAGTGTGCACCTCTCCATGGCATCTTCTGGCAA 1620
DB 1667 AGCCTCACATCAGTTTCTCTGTTGCAAGTGTGCACCTCTCCATGGCATCTTCTGGCAA 1726
QY 1621 TACTGGGAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1680
DB 1727 TACTGGGAACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGTGAAG 1786
QY 1681 TGTGGGAGCTTCCCTTTTATCTACTCTCTCATCAGCAGGCCCAATGGATACTCTCC 1740
DB 1787 TGTGGGAGCTTCCCTTTTATCTACTCTCTCATCAGCAGGCCCAATGGATACTCTCC 1846
QY 1741 CAATATGAATATTACCCCAACCAAGTAAAGTAAGCAATCAGGATTCGAAGTCTCTGGG 1800
DB 1847 CAATATGAATATTACCCCAACCAAGTAAAGTAAGCAATCAGGATTCGAAGTCTCTGGG 1906
QY 1801 CTTTATTTGCGAACCAATCCAGTGGAGTTCAATGTCTAGTCAATAGACAGAGATCA 1860
DB 1907 CTTTATTTGCGAACCAATCCAGTGGAGTTCAATGTCTAGTCAATAGACAGAGATCA 1966
QY 1861 CCTCAGTGACAAGAAAGTAAAGGAGCAGTGTCTCAGGGGGCAGAGAAATCAAGGGGTCC 1920
DB 1967 CCTCAGTGACAAGAAAGTAAAGGAGCAGTGTCTCAGGGGGCAGAGAAATCAAGGGGTCC 2026
QY 1921 TTTGGAAGCAAGGTTCATAAAAAATTTACTGCAGTTACTTACTCTGTCTCTGTATGACCG 1980
DB 2027 TTTGGAAGCAAGGTTCATAAAAAATTTACTGCAGTTACTTACTCTGTCTCTGTATGACCG 2086
QY 1981 GGGTCAATCTCTCTTGACCAATCCCCCTAGATTCAGTTGTAAGAAATCTTCTGTAG 2040
DB 2087 GGGTCAATCTCTCTTGACCAATCCCCCTAGATTCAGTTGTAAGAAATCTTCTGTAG 2146
QY 2041 TGTACACACCCCTCTGGAGTCTCTCTCTACATCTGGAGAGTATCTCTACATCCAA 2100
DB 2147 TGTACACACCCCTCTGGAGTCTCTCTCTACATCTGGAGAGTATCTCTACATCCAA 2206

QY 2101 TATGCATGGTCACTGTTTACAAGAGAACGACCCGATTTTGCAAGTTGCTGCAGATGG 2160
DB 2207 TATGCATGGTCACTGTTTACAAGAGAACGACCCGATTTTGCAAGTTGCTGCAGATGG 2266
QY 2161 GAATTCACACAGCTGAGGTAGCCAAAGATTACTGCACAGCCACTGGGAAAGACACAGCAG 2220
DB 2267 GAATTCACACAGCTGAGGTAGCCAAAGATTACTGCACAGCCACTGGGAAAGACACAGCAG 2326
QY 2221 TATTAACITCTTGTGGGACGGAATGTTGTCACAGGAGCAGCTAAGTCTCTAGAAGAA 2280
DB 2327 TATTAACITCTTGTGGGACGGAATGTTGTCACAGGAGCAGCTAAGTCTCTAGAAGAA 2386
QY 2281 GGAGAAATTAATGCATCTTCTAGATACCTGCTGCACAGGATGATCCTTAGTGTGACTCTC 2340
DB 2387 GGAGAAATTAATGCATCTTCTAGATACCTGCTGCACAGGATGATCCTTAGTGTGACTCTC 2446
QY 2341 TAAAGAACTACAGCCCCCAAGTGGAGAGTGGACAATAAATAGTCTAGTCCACAGCTC 2400
DB 2447 TAAAGAACTACAGCCCCCAAGTGGAGAGTGGACAATAAATAGTCTAGTCCACAGCTC 2506
QY 2401 CACCATTCCTAGCTCAAGTCAAGAGAAAGACCCCTAAAATTAAGACAGAGACAAGTGAAGA 2460
DB 2507 CACCATTCCTAGCTCAAGTCAAGAGAAAGACCCCTAAAATTAAGACAGAGACAAGTGAAGA 2566
QY 2461 GGGATCTGGAGACTTGGATAATCTAGATGCTTATCTTGGTGTGACTGACTGCTGACTT 2520
DB 2567 GGGATCTGGAGACTTGGATAATCTAGATGCTTATCTTGGTGTGACTGACTGCTGACTT 2626
QY 2521 TTACAATAAATTCATATCTCTCAATGGTGTAGTCTCTGGGGACTAAGCAACAGGTGTTTCA 2580
DB 2627 TTACAATAAATTCATATCTCTCAATGGTGTAGTCTCTGGGGACTAAGCAACAGGTGTTTCA 2686
QY 2581 AGGAACATAATCTCTGGGTTTGAAGTTTCAAGTCTGACGCTCTATCTCTCTCCATA 2640
DB 2687 AGGAACATAATCTCTGGGTTTGAAGTTTCAAGTCTGACGCTCTATCTCTCTCCATA 2746
QY 2641 TAAACGAGCAGTGTCTCTGGATAGCCCTGTTTCTTGTGGCTCAAGTCTCTCCAGTAAAAA 2700
DB 2747 TAAACGAGCAGTGTCTCTGGATAGCCCTGTTTCTTGTGGCTCAAGTCTCTCCAGTAAAAA 2806
QY 2701 TATCAGTGTCTTCCCATGTTTACCAAGCAACCCATGTTGGGTGGGAATCCAAAGATGAT 2760
DB 2807 TATCAGTGTCTTCCCATGTTTACCAAGCAACCCATGTTGGGTGGGAATCCAAAGATGAT 2866
QY 2761 GGATAGTCAGAGAAATATGCTCAAGTATGGGTGGGCCAAACCCAAATGTGACTGTGAC 2820
DB 2867 GGATAGTCAGAGAAATATGCTCAAGTATGGGTGGGCCAAACCCAAATGTGACTGTGAC 2926
QY 2821 TCAGACTCTTCTCCTCAGGAGACTGGGCTTACCAAACTTCAAAAGCCGCGAGAATGGAAAC 2880
DB 2927 TCAGACTCTTCTCCTCAGGAGACTGGGCTTACCAAACTTCAAAAGCCGCGAGAATGGAAAC 2986
QY 2881 TATGAATTCAAACTCCATGGGAAGACAGAGAGATTAATTAATCTTTTACCAGACC 2940
DB 2987 TATGAATTCAAACTCCATGGGAAGACAGAGAGATTAATTAATCTTTTACCAGACC 3046
QY 2941 TGCAGTGGGTGCTCTATTCCTCCATGCTCTCTGGTCTAATAGCATACAGGTGCGAG 3000
DB 3047 TGCAGTGGGTGCTCTATTCCTCCATGCTCTCTGGTCTAATAGCATACAGGTGCGAG 3106
QY 3001 ACCAGTATTGCAACAGCAGCAGCAGATGCTTCAATAGGCTGTGTGAATCCCCCATGGG 3060
DB 3107 ACCAGTATTGCAACAGCAGCAGCAGATGCTTCAATAGGCTGTGTGAATCCCCCATGGG 3166
QY 3061 AATGGGGCTAATCCCTATGCGCAAGCAGAGCATCTAACCACTGGGTTCCTGGCCCGA 3120
DB 3167 AATGGGGCTAATCCCTATGCGCAAGCAGAGCATCTAACCACTGGGTTCCTGGCCCGA 3226
QY 3121 TGGCATGTGTGCTGAGAACAGTTTCTCATGGCACACAAATAGGCCTCTTCTTAGGAA 3180
DB 3227 TGGCATGTGTGCTGAGAACAGTTTCTCATGGCACACAAATAGGCCTCTTCTTAGGAA 3286
QY 3181 TTCCCTGATGATCTTGTGGGCCACCTTCCAACTTGAAGGCCAGAGTGAACAAAGAGC 3240

[illegible]

Db 445 CTCCTTAGGACCGCTTTTACTACAGGCACTGGATGGTTTCTCTGTTGGTGAATCGAGA 504
Qy 478 GGCAACAACTGTATTTGTATCAGAAAAATGTACACAAATACCTGCAATATAACAGAGGA 537
Db 505 TGGAAACATTTGATTCGTGTGAGAAAAATGTACACAGTATCTGCAGTACAAGCAGAGGA 564
Qy 538 CCGTGTAAACACAGTGTTCACATATCTTACATGAAGAAGACAGAAAGGATTTCTTAA 597
Db 565 CCGTGTAAACACAGTGTTCACAGATCTTACATGAGCCAGCAAGGAAGGA-TTCTTAA 623
Qy 598 GAATTTACCAAAATACAGTAAATGGAGTTTCTGGACAAATGAGCCCCCAAGACAAAA 657
Db 624 ACACTTACCAAAATCCACAGTTAATGGAGTTCTTGGACTAATGAGAACAGACAAAA 683
Qy 658 AAGCCATACATTTAATTCGGGTATGT-TGATGAAGAACACACATGATATCTGGAGACA 716
Db 684 AAGCCCCATACATTTAATTTGCTGTATGTTGATGAAACACACACATTTTGAAGAGC 743
717 TAAAGCCAGTCCCTGAATGCGCCAGAGATATGAACAAATCAGTGTCTTGCCTGTCTC 776
744 TGAATGCCAGTCCCGAAACGCGCAGAGATATGAACAAATCAGTGTCTTGCCTGTCTC 803
Qy 777 AGCCACAGCTATGATGAGGAAGGGAGATTTGCAATCTGTATGATCTGTGTGGCAC 836
Db 804 AGCCTCGGCTATCTGGAAGAGAGAGACTTGCAGTGTGTATGATCTGCTGGTGC 863
Qy 837 GCGCATTTACTACAGGAGAGAAACATTTCCATCAAAACCTCGAGAGCTTTATTACCAGAC 896
Db 864 GCGCGTGACTGC-----GCCATCCCATCCAGTCCGAGAGCTTTATTACCAGAC 914
Qy 897 ATGATCTTTAGGAAGTGTTCATATAGATACAAATCAGTGTCTTGCCTGTCTC 956
Db 915 ATGACCTTTCCGGAAGTGTTCATATAGATACAAATCAGTGTCTTGCCTGTCTC 974
Qy 957 CTGGCTTTGAAGATATAATCCGAGGCTGATTCAGAGATTTTGTAGTCTAAATGATGGC 1016
Db 975 CTGGCTTTGAAGACATATCCGAGAGTATCCAGAGGTTCTTCTGATGATGGC 1034
Qy 1017 AGTCATGTCGCCAGAACGTCCTATCAAGAGCTTATCTTAATGGCCATCGAGAAACCC 1076
Db 1035 AGTCATGTCGCCAGAACGTCCTATCAAGAGCTTATCTTAATGGCCATCGAGAACCC 1094
Qy 1077 CAGTATATCGATCTCGTTGGCTGATGAACTATAGTACTGCACAGACAAAAAGCAAC 1136
Db 1095 CCGTGTATCGTTTCTCTGCTGATGAACTATGAGTGGCGGACACAAAAAGCAAC 1154
1137 TCTTCCGAAATCCTGTAACAAATGATCCAGATGGCTTGTCTCAACCCACTTCTTCA 1196
1155 TCTTCCGCAATCCTGTAACGAATGATCTCAGGGCTTCATCTCGACCCACTTCTTCA 1214
Qy 1197 GAGAACAGATGGATATAGACAAACCCAAATCCTGTTGGACAGGGATTAGACCAACCTA 1256
Db 1215 GAGAACAGATGGATATAGACAAACCCAAATCCTGTTGGACAGGGATTAGACCAACCTA 1274
Qy 1257 TGGCTGGATGCAACAGTTCCGTTAGCGGCGATGAGTATGTGCGCAACCAAGCTTTACAGA 1316
Db 1275 CAGCAGGGT-----GTGGCGTGAGCATGTCTCCAAATCAGAATGTACAGA 1319
Qy 1317 TGCCGACGACGAGGCGCTATGGCTGGCAGACCTTAGCACCACAGGGCAGATGAGTGAG 1376
Db 1320 TGATGGGACGCGGACCTATGGCTGGCAGACCTTAGCACCACAGGGCAGATGAGTGAG 1379
Qy 1377 CTAGTATGGGGTTTCCAGTAACATAGCTTTCATGACCCCTGGGCGCAGGCATGCAATCAC 1436
Db 1380 CTAGTACGGGGTTTCTAGTAGGCTAGCTTCACTGACGCCAGGACAAAGCCTACAGTCG 1439
Qy 1437 CATCTTCCTACAGAACAACTATGGGCTCAACATGAGTAGCCCCCACCATGGGAGTC 1496
Db 1440 CATCTTCCTATCAGAACAGCAGTATGGCTCAGCATGAGCAGTCCCCCCCCACGCGAGTC 1499
Qy 1497 CTGGTCTTGGCCCCAACACAGCAGATATCATGATTTCTCTCGTAAATCGTGGGAGTC 1556
Db 1500 CTGGTCTTGGCCCCAACACAGCAGATATCATGATTTCTCTCGTAAATCGTGGGAGTC 1559

Qy 1557 AGATAGCCTCACATCAGTFTTCTCCTGTGCAAGGTGCACTCTCCCATGCACTTTCTG 1616
Db 1560 AGATGGCTCCACACAGTCTCTCTGCTGCAAGGTGCACTCTCCCATGCACTTTCTG 1619
Qy 1617 GCAATCTGGGAACACACAGCTTTTCCAGCAGCTCTCTCAGTGGCCCTGCAAGCCATCAGT 1676
Db 1620 GCAACACAGGAGGACACAGCTTTTCTAGCAGTCCCTCAGTGGCTTGCAGCCATCAGT 1679
Qy 1677 AAGGTGTGGGACCTTCCCTTTTATCTGTCTCATCCACAGGCCCCAAATTTGATAACT 1736
Db 1680 AAGCGTGGGACCTCTCTTTTATCTGTCTCTCACCAGGCCCCAAATTTGATAACT 1739
Qy 1737 CTCCCAATATGAATATTAACCAACAGTAAGTAAGCAATCAGATTTCCAAAGTCTCTC 1796
Db 1740 CTCCCAATATGAATATTAAGCCAGCAAGTAAGTGAAGTGGTTCAGGACTCTAAGAGCCCC 1799
Qy 1797 TGGGCTTTTATGGACCAAAATCCAGTGGAGAGTTCATGTCAGTCAAAATAGCAGAG 1856
Db 1800 TAGGCTTATCTGTGAACAGAAATCCAGTGGAGAGTTCAGTGTGTCAGTCAAAACAGCAGAG 1859
Qy 1857 ATCACTCAGTGCACAAAGAAAGTAAGGAGAGCAGTGTGAGGGGCGACAGAAATCAAGGG 1916
Db 1860 ATCCCAAGTCAAAAAGAAAGAAAGAGAGCAGTGGGAGGTGTGACAGAGAGCCAGGG 1919
Qy 1917 GTCTTTTGGAAAGCAAGGTCTATAAAATTAATCTGCACTTACTTACCTGTCTTCTGATG 1976
Db 1920 GACCTCTGGAAGCAAGAGGCCCAAGAAATGCTGCACTTACTCAGTGTCTCTCCGACG 1979
Qy 1977 ACCGGGTCTATCTCTTGGACCAACTCCCCCTTAGATTCAGTGTAAAATTTCTG 2036
Db 1980 ACCGAGGCAATCTCTCTTGGACCAACTCCCCCTTAGATTCAGTGTAAAATTTCTG 2039
Qy 2037 TTAGTGTCAACAGCCCTCTCTGAGTCTCTCTCTACATCTGGAGAGTACTCTCTACAT 2096
Db 2040 TTAGTGTCAACAGCCCTCTCTGAGTCTCTCTCTCAACATCAGGAGCAGTGTCTCCACT 2099
Qy 2097 CCAATATGATGCTGCTACTTTACAAGAGAACGCCGATTTTGCACAAAGTGTGTCAGA 2156
Db 2100 CCAATATGATGCTGCTACTTTACAAGAGAACGCCGATTTTGCACAAAGTGTGTCAGA 2159
Qy 2157 ATGGAATTCACAGCTGAGTACCAAGATTAAGTGCACAAAGCCTCTGGGAAGACACA 2216
Db 2160 ATGCAACTCTCCCAAGGAGTTCGCAAGATCACTGCAAGAGGCCACTGGGAAGACACA 2219
Qy 2217 GCAGTATAACTTCTTGGGACGGAATTTGTCAACGAGGAGCAGTAAAGTCTCTAAGA 2276
Db 2220 GCAGACTGCTTCTGTGGAGAGGGA---CAACAGCAGGAGGAGCTGAGTCTCTAAGA 2276
Qy 2277 AGAAGGAATTAATGCACTTCTTAGATACCTGCTGGACAGGAGTATCTTAGTATGAC 2336
Db 2277 AGAAGGAATTAATGCTCTGCTTAGATACCTGCTGGACAGGAGTATGACCTAGTATG 2336
Qy 2337 TCTCTAAGAACTACAGCCCCCAAGTGGAGAGTGGACAAATAAATGAGTCAAGTACACA 2396
Db 2337 TTGCAAAAGAGCTGCAAGCCCCCAGGCGCAGTGGGAGACAGTAAATGAGTCAAGT 2396
Qy 2397 GCTCCACCATCTCTAGCTCAAGTCAAGAGAAAGACCTTAATTAAGACAGACAGCAAGT 2456
Db 2397 GCTCCACCAATCCCAAGCTCTGGCCAAAGAAAGACCCCAAAATTAAGACAGGAGAAC 2456
Qy 2457 AAGAGGATCTGGAGACTTGGATATCTAGATGCTATTTCTTGGGATCTGACTAGTCTG 2516
Db 2457 ACAGGATCTGGAGACTTGGATATCTAGATGCTATTTCTTGGAGATTTGACAGTCTG 2516
Qy 2517 ACTTTTACAATAATTCATATCTCAATAGTATCTATCTGGGAGTCAAGCAACAGGAGT 2576
Db 2517 ACTTCTACAATAATCTA-----CAATGGGCTGACCCAGGCGCCAAACAGCAGATGT 2570
Qy 2577 TTCAAGGAACTAATCTCTGGGTTTGAAGATTCACAGTGTGCACTGTCTTCTGCTCTC 2636
Db 2571 TTGAGGAGGAGGTTCTCTGGGTTTGGAGTCCACAGCCTGTGCACTGTCTGCTCTC 2630

AUTHORS Kim,H.-J., Lee,S.-K., Na,S.-Y., Choi,H.-S. and Lee,J.W.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1998) Pharmaceutical Sciences, Chonnam National
Univ., 300 Tongpung-dong Puk-ku, Kwangju 500-757, Korea
FEATURES
source 1. .4546
/organism="Xenopus laevis"
/db_xref="taxon:8355"
CDS
1. .4176
/note="member of the SRC-1 superfamily of transcription
coactivators"
/codon_start=1
/product="retinoid X receptor-interacting coactivator
XSRC-3"
/protein_id="AAC12927.1"
/db_xref="PID:g2852394"
/db_xref="GI:2852394"
translation="MSGIGENSLDPLASETRKRPSSCDTPGCLTSGEKKRREQES
KYLEELADLISANLSDIDNENVPDKCAILKEIVQRIQKEOGKASNNDDVQKADY
SITGGQVIDKDSGLPLQALDGLFVYNNREGSIVFSENVTLQIKQEDLVNTSVI
SILHEERDKDFLKNLPKSTYNGVPWFSETPRQKSHTFNCRMLVKTSHDLEDLSNDA
KQVETMOCLASOPRAMIEGEDLOCMICVARRITTAERAFSANPESITRDLITG
KVNIDANLSRSMRPGFEDTIRRCIORFLFHSEGOPTVKRYHOEAYVHLSSTPL
RFLSLADCTMTAOTKSLFRNPVNDPHGFVSTHFLQREONGVDPNPNAQGTROP
NPPLPNTMNSPQAOQQONRNNGMDPNMAQOMQRYKSPGNMAVYNGAPGVQOSP
IQNNNIGLNNPFPHGSPGMNANQNLVSPRNASNOFSPVGMNPMGSS
GNAGGFSFSSLSALHAISEGVSSLSLSPGQVKNNSNMNPMPOQKICNQDK
SPGLYCEQGVQSVQSGREHLGCEKVENIFEGESQROAESKGHKLIQLIT
CFTEERGOSLSSMDCKDSNVTSFSGVSSSTSIGVSTSLHGSMLQEKHRLHK
LLONGNSPAEVAKITAEATKDVQETVSSAPCTEATVKEQLSPKPKENALLRH
DKDDKPLADKIDPKVEHMDIKMGSSSSNVTSFSSQDKVKIKTERGEVPGDNL
DALLDGADGDFYSNSRASDLGPKVPQDPSQTLAMRSPOSMSGRPPFNAMSL
DSRSSTPVRNVNSFPMPLPKMGQDPMQDQDFGVGMGSPGRNMGPPGDDWAL
QNSAVNLEPNVSGVPGPDYSAMTRPAGMMPGLLTRNSIPGSRPVMOQOQH
ILPMDNDMAMSGSNYPGOQAPSNPPGWPDAIMNMQGRGAQNRQLGRNSLDLLC
PPSTVEGQTEIALLDLHLLNSTNDATGLIEDRALGIPOLYSOGALPQDPSTOP
QGSVPIDQRPYQHYAGGAMSAGGNMGQHPPEYVGMQOQGHPLPG
MHPNANLIRPNPILKQRLMQLQORLQGLQNLNOLKEMKVDMPNPGGAVPRVYM
QTPVSGGFLNAQVAQKRELISHQIRHMMAMMOQOQPOAFSPPNVTSASAM
DNPLGPPMPQAPQPSYPNTYNGINQOTDPTFGRVSSPPNMMSSRMASQNPHTOT
TOMYPSDPMKGWPSGNMARPNRPQOQYSHQTNPATYNNMHNMGNGHMGQMINSLP
MSGMPMGPDQKCYC"
BASE COUNT 1309 a 1150 g 957 t
ORIGIN

Query Match 39.9%; Score 1793.4; DB 4; Length 4546;
Best Local Similarity 66.7%; Pred. No. 0;
Matches 2856; Conservative 0; Mismatches 1291; Indels 138; Gaps 15;

Qy 86 ATGAGTGGATTAGGAGAAAA---CTTGGATCCACTGGCCAGTGATTCACGAAAACGCCAA- 142
Db 1 ATGAGTGGATTAGGAGAAAACTCTTGGATCCGCTGGCTCGAAACAAAGAGCGCAAG 60

Qy 142 --ATTGCCATGTGATACTCAGGACAGAGTCTTACCTGCAGTGGTGAAGAACGGAGCGG 199
Db 61 CCGTCTTCTGTGATACCCCGGACCTGGTCTGACCTGAGTGGTGAAGAGAGGCGGT 120

Qy 200 GAGCAGGAAAGTAAATATATTGAAGAAATGGCTGAGCTGATATCGCAATCTTAGTGAT 259
Db 121 GAGCAGGAGAGCAAGTACATAGAGAGTGGCCGATCTCATTTACGCCAATCTTAGCGAC 180

Qy 260 ATTGACAATTTCAATGTCAAAACAGATAAATGTGCGATTTTAAAGGAAACAGTAAGACAG 319
Db 181 ATAGACAATTTCAATGTCAAAACAGATAAATGTGCGATTTTAAAGGAGAGCGGTACGGCAG 240

Qy 320 ATACGTCAAAATAAGACAGAGAAAACTATTTCCTCATGATGATGATGTTTCAAAAAGCC 379
Db 241 ATCCCTCAGATTAAGAGAGAGAAAAAGCCCTCTTCTAATGACGAGAGTGTGCAGAAAGCC 300

Qy 380 GATGATATCTTACAGGCGGAGGAGTATTATGATAAGAACTCTTAGGACCGCTTTACTTT 439
Db 301 GATGTTCTCTACAGGCGGAGGAGTATTATGATAAGAACTCTTAGGACCGCTTTACTTTG 360

Qy 440 CAGGCATTGGATGGTTTCCCTATTGTTGGTGAATCGAGAGCAACACATGTTATTGTATCA 499
Db 361 CAGGCTCTGGATGGCTTTCTGTATGTGTAATCGAGAGGCGAGCATGTTATTGTATCT 420

Qy 500 GAAATGTCACACATACCTGCAATATAAGCAAGAGACCTGTTACACAAGTGTATTAC 559
Db 421 GAAATGTCACCCAGTACTTACAATAACAACAGAGGATCTGGTTAAACACTAGTGTCTAC 480

Qy 560 AATATCTTACATGAAGAGACAGAAAGGATTTCTTAAAGATTTTACCAAAATCTACAGTT 619
Db 481 AGCATTCTGCACGAGGAGGACCGTAAGATTTCTCAAGAACCTGCCNAATCTACTGTT 540

Qy 620 AATGGAGTTTCTCGACAAATGAGCCCCAAAGACAAAAGCCATACATTTAATTGGCGGT 679
Db 541 AATGGCGTCCCATGGTTTAGTGAACCCCTCGACAGAAGAGCCACACGTTCAACTGCGGC 600

Qy 680 ATGTTGATGAAGAACACCATGATATTCTGGAACACATAAAGCCAGTCTCTGAATGGCG 739
Db 601 ATGCTGGTCAAAACCTCTCATGATCATCTGGAGGACGGAA---GCACCTGGAGCGCGGT 657

Qy 740 CAGAGATATGAACAAATGACAGTCTTGGCCCTGCTCAGCCACGAGCTATGATGGAGAA 799
Db 658 CAGAGATACGAGACCATGACAGTCTGCTTTTATCTCAGCCCCGGCGCATGATCGAGAG 717

Qy 800 GGGGAAGATTTGCAATCTTGTATGATCTGTGTGCGCCGCGCATTTACTTACAGGAGAA 859
Db 718 GGAGAAGACCTTCAGTCTCTGTATGATCTGTGTGCGCAGGCGCATCACAACTGCAGAGG 777

Qy 860 ACATTTCCATCAAAACCTTGAGAGCTTTTATACGACACATGATCTTTCAGGAAAAGTTGTC 919
Db 778 GCCTTTCAGCCAATCTGAAAGTTTCATCAAGGCGACGATCTGACCGGCAAGGTGGTG 837

Qy 920 AATATAGATACAAATCTCACTCAGATCTCTCCATGAGGCTGCTTTGAAGATATTAATCGA 979
Db 838 AACATTGATGCAACTCCCTCGCTTCTCCATGAGACGGGCTTTGAAGATACCATCGT 897

Qy 980 AGGTGATTCAGAGATTTTGTAGTCTAAATGATGGGAGTATGTTCCCGAAGAACGTGAC 1039
Db 898 AGATGTATTCAAAGTTCTCTGTTTACAGCGGAAGGCGACGATGACCTACAAACGCCAT 957

Qy 1040 TATCAAGAGCTTATCTTAATGGCCATGCGAAGAACCCAGTATATCGATTTCTGTTGGCT 1099
Db 958 TACCAAGAGGCTATGTCACGAGCTTTCTGAAACGCTCTGTATAGTCTTCTCTAGCA 1017

Qy 1100 GATGGAATCTAGTGACTGACACAGACAAAAGCAAACTCTTCGAAATCTCTGTAAACAAT 1159
Db 1018 GATGGCACCATTGGTACTGCTCAGACAAAAGTAAGCTCTTCCGGAATCTCTGTACCCT 1077

Qy 1160 GATCGACATGGCTTTGTCTCAACCCACTTCTTCAGAGAGAACAGATGATATAGACCA 1219
Db 1078 GATCGGACGCTTTGCTCCACGCAATTTCTTGCAAGGGAACAGATGGCTACCGACCC 1137

Qy 1220 AACCAATCTGTTGGACAGGATTAAGACCCTATGCGCTGGATGCAACAGTTCGGTA 1279
Db 1138 AATCCGAACCAATGGCTCAAGGTATTGCGCCACAAATG-----AACCCCTAACCTA 1188

Qy 1280 GGCAGCATGATGTGCGCCAAACCAAGCTTACAGATCGCGAGCAGCAGGCGCTATGGC 1339
Db 1189 CCCAATCAATGAATTAATGCCAGCGCCATGACAGCAGAGAACAGTATGGC 1248

Qy 1340 TTGGCAGACCTTAGCACACACAGGCGCATGATGAGTGGAGTATGCGGGTTCCAGTAAAC 1399
Db 1249 ATGGGGACCCCAACAGCATGCGACAAATGCGAGGCGATGATACAACTCCCCAGGCAAC 1308

Qy 1400 ATAGCTTATTGACCCCTGGGCGAGGATGCAATCACCCTCTTCTACCGAAGAACCAAC 1459
Db 1309 ATGCGACCTGTTAATCAGCTCCAGGGGTACAGAGTCTCGGTACCAAGAACATAGTAAT 1368

Qy 1460 TATGGCTCAACATGATAGCCCCCAGTCTGCTGCTTGTGCCCCAACCAAGCAG 1519
Db 1369 TATGGCTTAAACATGAACAGCCCGCCAGTCCGGGCGATGAATGCGCAACCAACCC 1428

Db 3559 ATGAAGTGGACCAATGAATCAGGTGGAGCTGGAGTATGAGGCCCGTGTGCAACA 3618
QY 3718 -----GCCCCAGCAGGGTTTCTTAATGCTCAATGTCGCCCAACGCGCAGCAGAGCTG 3772
Db 3619 CCAGTCTCAGCAGCAGGGCTTCTCAACGCGCAGATGTCGCGCAGAGAACAGGGAAGT 3678
QY 3773 CTAAAGTCATCCTCCGACACAGAGGGTGGCTATGATGATGTCGAGCAGCAGCAGCAG 3832
Db 3679 ATTAGCCACCAATACAGACAGCAAGATGGCCATGATGATGATGATGATGATGATGAT 3738
QY 3833 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3892
Db 3739 C-----CTCAA 3744
QY 3893 GCCTCAGCCCACTCTTAATGATGATGCTTCCCGCAGCAGATGATGATGATGATGATGAT 3952
Db 3745 GCCTCAGCCCTCTCCCAATGATGATGCTTCCCGCAGCAGATGATGATGATGATGATGAT 3804
QY 3953 CCACATGTCACCAAGCTCTCCGACACAGTTCCTATATCAACCAATATGATGATGATGAT 4012
Db 3805 CCACATGTCACCAAGCTCTCCGACACAGTTCCTATATCAACCAATATGATGATGATGAT 3864
QY 4013 CAACACACAGATCAGCTTTGCTGAGTCTAGTCTCTCCCAATGCAATGATGATGATGAT 4072
Db 3865 CAGCAGATGACCCACTTTTGAAGGGTCTAGCCCACTATGATGATGATGATGATGATGAT 3924
QY 4073 AGAATGGTCTCCCAAGATCCCATGATGATCAACACCGCAGCTGATCCTATCATCAG 4132
Db 3925 CGATGCTCTCTCCCAAGATC-----TCACCTCAGCAGCAGCAGCAGCAGCAGCAGCAG 3975
QY 4133 TCCTCAGAAATGAAGGCTGGCCATCAGGAAATTTGCCAGGACAGCTCTTTTCCAG 4192
Db 3976 TCACAGATCAAGGCTGGCCATCAGGAAATTTGCCAGGACAGCTCTTTTCCAG 4035
QY 4193 CAGCAGTTTCCCAACAGGGAATCTGCAATGATGATGATGATGATGATGATGATGATGAT 4252
Db 4036 CAGCAGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4095
QY 4253 AGTGGTCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 4312
Db 4096 GCGAATCAGATGCGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4155
QY 4313 CCGTATCAGAAATGCTGCTGATC 4337
Db 4156 CCGTATCAGAAATGCTGATGATC 4180

MULT 7

LOCUS HS1049G16 86453 bp DNA PRI 06-APR-1999
DEFINITION Human DNA sequence from clone 1049G16 on chromosome 20q12-13.2
Contains gene similar to GLUCOSAMINE-6-SULFATASE, a nuclear
receptor coactivator gene, ESTs, STSS, GSSs, complete sequence.
ACCESSION AL034418
NID 94455409
VERSION AL034418.2 GI:4455409
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 86453)
Direct Submission
Submitted (06-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 21, 1999 this sequence version replaced gi:4375895.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
This sequence is the entire insert of clone 1049G16. The true left
end of clone dj237J2 is at 54555 in this sequence. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
1049G16 is from the library RPC15 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: PCYPAC2.

FEATURES
Location/Qualifiers
1..86453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="1049G16"
/map="q12-13.2"
/clone_lib="RPC15"
4..293
repeat_region
/note="AluX repeat: matches 4..296 of consensus"
949..1162
repeat_region
/note="MER20 repeat: matches 9..218 of consensus"
1495..1657
repeat_region
/note="I2 repeat: matches 2300..2466 of consensus"
2214..2268
repeat_region
/note="MER58 repeat: matches 1..56 of consensus"
3333..3737
repeat_region
/note="L1MB5 repeat: matches 5765..6176 of consensus"
4004..8719
gene
/gene="dj1049G16.1"
CDS
join<4004..4232,4539..4635,5199..5293,6870..7039,
7280..7422,8596..8719)
/gene="dj1049G16.1"
/note="match: multiple ESTs; match: AA612659 AA323130
AA294950 AA332792 R14440 W71227 AA304723 AA506790 N39498
AA138508 AA120567 AA727360 C00596 AA295197 AA612659"
/codon_start=3
/evidence=not_experimental
/product="dj1049G16.1 (similar to
GLUCOSAMINE-6-SULFATASE)"
/protein_id="CA840661.1"
/db_xref="PID:el424046"
/db_xref="PID:94584217"
/db_xref="GI:4584217"
repeat_region
5662..5835
/note="MER3 repeat: matches 9..209 of consensus"
5836..5866
repeat_region
/note="MER3 repeat: matches 18..47 of consensus"
7598..7639
repeat_region
/note="2 copies 21 mer 100% conserved"
9263..9559
repeat_region
/note="I2 repeat: matches 2000..2368 of consensus"
9723..9881
repeat_region
/note="FRAM repeat: matches 3..161 of consensus"
complement(9968..10276)
misc_feature
/note="match: GSS AQ233123"
10229..10324
repeat_region
/note="MIR repeat: matches 120..236 of consensus"
prim_transcript
<10761..>11097
/note="match: multiple ESTs; match: AA032808 AA059962
AA138552 A1046944 A1197066 C00596 W50675 W53630 W54104

```
misc_feature
repeat_region
mRNA
W58927 W97878
<12013..12427
/note="match: GSS AQ035209 clone 2329M4"
13581..13604
/note="12 copies 2 mer aa 100% conserved"
complement(join(14705..17086,17420..17561,17912..18086,
19216..19510,21395..21487,23126..23419,28108..28279,
30441..30567,30670..30870,31290..31484,32709..32844,
33730..34601,34779..35170,36297..36474,36856..36996,
42459..42570,42743..42931,43316..43490,45011..45111,
46409..46581,48162..48265))
/note="dJ1049G16.2"
/note="match: CDNAS AF016031 AF012108 AF036892 AF010227
AF000581 AF044080 X97674 U39060 U80737 U64828 U64606
U56920 U59302 AJ000881 AJ000882 M60725 AJ005690; match:
ESTs AA0265268 AA488485 AA114092 AA065270 W84822 R28559
AA085272 AA300819 T77368 AA065271 W84775 AA045419 AA150333
R21765 R66357 AA065269 W46433 AA384019 AA764263 AA811653
AA922415 AA828040 AA905298 AA729022 AA360136 R21718 W46488
H52110 AA678029 AA045313 H51992 AA767048 R67499 W20355
C05232 AA665470 AT146974 AA229716 AA230072 AA305762
AA94062 AA346492 AA346493 AA809279 AA738120 AL270423
AA488334 AA678419 AA883388 AA824532"
evidence=not_experimental
complement(14705..48265)
/note="dJ1049G16.2"
15605..15642
/note="19 copies 2 mer ac 95% conserved"
15732..15921
/note="dJ1049G16.2"
/note="match: STS G42545"
16761..17000
/note="dJ1049G16.2"
/note="match: STS G15161"
complement(join(17075..17086,17420..17561,17912..18086,
19216..19510,21395..21487,23126..23419,28108..28279,
30441..30567,30670..30870,31290..31484,32709..32844,
33730..34601,34779..35170,36297..36474,36856..36996,
42459..42570,42743..42931,43316..43490,45011..45111,
46409..46581,48162..48244))
/note="dJ1049G16.2"
/note="match: O09001"
/codon_start=1
evidence=not_experimental
/product="dJ1049G16.2 (nuclear receptor coactivator)"
/protein_id="CAB40662.1"
/db_xref="PID:g4584218"
/db_xref="GI:4584218"
/translation="MSGIGENLPLASDRKRKRLCDTPGGLTSCGEKKRRREQSKY
IEELAEISANLSDIDNFNVKDKALIKETVROIKEQKGTISNDDVDQKADVS
TGQVIDKDSGLPLLQALDGLFVNRDGNIVSENVQYLOKVEDLVNYSYNI
LHEDRKDFLKNLPKSTVGVSTNETQKSHFTFNCMLMKTDPHILEDLVNINSPMR
QRYETMQCFALQSPRAMEGEDLQSCICVARRITGERTFPSPNPFSTRHDLGK
VYNTDTSLSRSMRPFEDIIRRCIQRFSLNDQGSQKXHYQEVSTSDGIFSPATL
NGHATVPYRESLADGIVTAOTSKLFRNPVTNDRGFVSTHFLOEQNGYRNPNP
VGOCIRPDMACNSVSGMSNPNQGLQMPSSRAYGLADPSTTGOMSGRYGSSNLA
SLTGPQMSQSPSSQNNYGLNMSPPHGPSPLAPNQNMISPRNRGSKIASHQFS
PVAGVSPMSGGTGNHSSSSLSALQALISEGVTSLJLSTSSPGPKLNDSPNMI
TOPSKVNDQDSKPLGYCDQNPVSESSMCQNSRDHLSDEKSSVEGAEHQGLPE
SKGHKLILQITCSDDRGSHSLNPLDSCKSSSVTSFGVSSSTGGVSSSTL
MHGSLQEKHRLILHLQNGNSPAEVAIKTAEATGKDTSSITSCGDVNVYKQGLSPK
KKENALLRYLLDRDPSDALSKELQVQEVGDNKMSQCTSSITPSSSEKDKPIKTE
TSRSGSDLDNLDAILDGLTSSDFYNNISISGNSLGTQKQVFGTNSLGSQSVQ
SIRPPYNRAVLSFVSVGSSPVKYNISAFMLPKQPMGLGNPRMDSQENYSSSQD
WQPNKAGRPNNSMGRPGDYNTSLPRALGSIPTPLRSLNISIPGARVLQO
QQQLQMRPGIPIPMGANPYGQAAANQLGSPDGMLSMEQVSHGTQNRPLRSLD
LGVGPNLEQSDERALLDQLTLLSNTDGLTLEEDRALGIPELYNQGLRPFQD
AFQGEAAVMDQKGLYQOTYPAQPGMGGFHLQGSFNSNMNMQOQNFQ
GMHPRANIMRPTWPKLRMLQLRQGLQFNLQSRQALEKMENTAGCAVRRPM
MQPQOQELNAQVRSRELLSHHFROORVAMMMQQQQQQQQQQQQQQQQQQQQ
QQQQQAFSPPPNTPASPSMDHLGATPMPAPQPPQPPYQNTGMDGPPAFGRYS
SPPNAMSSRRGSPQNPMQHPQAASIYQSSEMKGWPSPGNLARNSSFSQQQFAHQGNP
```

```
repeat_region
repeat_region
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
misc_feature
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
misc_feature
misc_feature
repeat_region
repeat_region
repeat_region
repeat_region
Query Match 24.1% Score 1084.2; DB 10; Length 86453;
Best Local Similarity 86.5%; Pred. No. 5.1e-249;
Matches 1285; Conservative 0; Mismatches 23; Indels 177; Gaps 1;
Qy 1187 TTCCTTCAGAGAACAGATGATAGACCAACCAATCCTGTGGACAGGGATT 1246
|||
Db 35181 TTGTTTTCAGAGAACAGATGATAGACCAACCAATCCTGTGGACAGGGATT 35122
Qy 1247 AGACCACTATGGCTGGATGCAACAGTTCGGTAGCGGCATGATGTCCGCCAACCA 1306
|||
Db 35121 AGACCACTATGGCTGGATGCAACAGTTCGGTAGCGGCATGATGTCCGCCAACCA 35062
Qy 1307 GGCTTACAGATGCCAGCAGCAGGCCTATGGCTTGCAGACCCCTAGCACCACAGGGCAG 1366
|||
Db 35061 GGCTTACAGATGCCAGCAGCAGGCCTATGGCTTGCAGACCCCTAGCACCACAGGGCAG 35002
Qy 1367 ATGAGTGGAGCTAGGTATGGGGGTTCAGTAACATAGCTTCATTGACCCCTGGGCCAGGC 1426
|||
Db 35001 ATGAGTGGAGCTAGGTATGGGGGTTCAGTAACATAGCTTCATTGACCCCTGGGCCAGGC 34942
Qy 1427 ATGCAATCACCATTCTCTACCAGAACCAACACTATGGCTCAACATGAGTAGGCCCCCA 1486
|||
Db 34941 ATGCAATCACCATTCTCTACCAGAACCAACACTATGGCTCAACATGAGTAGGCCCCCA 34882
Qy 1487 CATGGAGTCTCTGCTTGCCTCCCAACACAGCAGAAATCATGATTTCTCCTCGTAATCGT 1546
|||
Db 34881 CATGGAGTCTCTGCTTGCCTCCCAACACAGCAGAAATCATGATTTCTCCTCGTAATCGT 34822
Qy 1547 GGGAGTCCAAAGATAGCTCACATCAGTTTCTCCTCTGTG----- 1587
|||
Db 34821 GGGAGTCCAAAGATAGCTCACATCAGTTTCTCCTCTGTGAGGATTTTGTGTTGACATT 34762
Qy 1587 ----- 1587
|||
Db 34761 TCCTTTTATTTTCTTTTAAAGTAATATTCTTCCATACTACTATAATTTCTGTAA 34702
Qy 1587 ----- 1587
|||
Db 34701 AGTTAATCTATTTTATATAGAAAATTTGAAGAAATTAAGTTGTATTTTGTACAGTGTCT 34642
Qy 1587 -----CAGGTGTGCACTCTCCCATGCA 1609
|||
```


	alternatively spliced; nuclear receptor coactivator; TIF2 gene; transcriptional mediator.
human.	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
1 (bases 1 to 6156)	
Voegel,J.J., Heine,M.J., Zechel,C., Chambon,P. and Gronemeyer,H.	
TIF2, a 160 kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors	
EMBO J. 15 (14), 3667-3675 (1996)	
96312984	
2 (bases 1 to 6156)	
Voegel,J.J.	
Direct Submission	
Submitted (22-APR-1996) J.J. Voegel, IGMB Inst.de Genet.et Biol.Mol.et Cell., CNRS-INSERM-univ.Louis Pasteur, B.P.163, C.U. de Strasbourg, F-67404 ILLKIRCH CEDEX, FRANCE	
Revised by author 25-JUL-96 and 10-MAR-97	
On Mar 11, 1997 this sequence version replaced gi:1490315.	
Related sequences U39060, U40396.	
Location/Qualifiers	
1..6156	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone_lib="lambdaExlox Ref.No.56"	
/tissue_type="placenta"	
163..4557	
/gene="TIF2"	
/note="TIF2"	
/function="transcriptional mediator for ligand-dependent activation function AF-2 of nuclear receptors"	
/codon_start=1	
/product="transcriptional intermediary factor 2"	
/protein_id="CAA66263.1"	
/db_xref="PID:e307031"	
/db_xref="PID:g1877215"	
/db_xref="GI:1877215"	
/db_xref="SPTREML:Q15596"	
/translation="MSGMGENTSDPSRAETRRKKCEPDQLGSPFKRNTEKRNRQENK YIEELAEIIFANFDIDNFNPKDKCAIKETVKQRIKEQEKAANAINDVEQKSD SSTGGVIDKDALGPMLLEADLGFVFNLEGNVFVSENVTOYLRYNBELNMKVSY SILHVGDHTFEVKNLPKLSIVNGSGWSEPPRRNSHTFNCRLMKVLPLDSEESGHNDI EAHQKYETMQCFAYSQSKSEGEDLOSLCICVARYPMKRPVLPSSSFTRDL QGKTSLDTSTWRAMPGWEDLVRRCKRFQAHEGESVSYAKRHHEVLRGLAFS QIYRFSLDGTFLVAATKSILRSQTNEPVLISHLHREQNVCMPNDLTGQTGM KPLNPISNSPAHQALCSGNPGDMTLSSINFPINGPKQMGPMPRGFGSGGMNHV SGMAQTPQGSNYALKNMSQSPGSMNPQOPTSMLSPRRHSPGVAGSPRIPOFS PAGSILHPVGVCYSTGNSHSYTNSLNALOALSEGHCSTGSSLASGLDLMKWLONS VNMAPPILSKMGSIDSKDCFLGEGPESETTGGAESCCHGEQKENDPNLPPAVSE RADQSOLDSKGTKLQLLTNTKSDMEPSPLASSLSDTNKDSTGLSPGSGTHGTS LKEKHKLHLRLDDSPVDLAKTAETAETKDLQSSSSSTAPGEVTIKOEPVSPKK ENALRYLLDRKDDKIDLPEITPKLERLDSKTDPASNTKLIAMKTEKEMSPEPDQ PGSELNLEEITDLLONSQLPOLFPDFPRCPAGASVDKQAIINDLMQTAENSPTVP GAKTLARIQSOTENPNRPGQLGLLEPNQLPIDITLQSPTAGPFPPIENSSPYVI PQPMGMNQHMIGNQNLGNSVTGMIGNSASRPTMFSGEWAPOSSAVRVTCATTASM NRPOVGMIRNPAASIPMRPSSQGRTLOSQVMNIGPSELEMNNMGPOYSQQQPH NOTAFWESPFIIDQAFSAQRNPQFGSPDDLCLCPHAASESPDEGCALLDYLYLAR NFDGLEIDRALGITPELVYSQAVDPEQFSQSDSNMLEQKPAFLQLOHRAQOQNRQ SYSPMDPFHTMGVRPSYATLRMPRGLRPTGLVNQNPOLRLOLQHRLOAQOQNR PLMNQISNVSNVLTLRPVPTQAPINAQLAORQEILLNOHLRQRMHQOOVOQRT LMQRGGLNTPSWVAFSGMPATHSNPRIFOAQAOPFFPFGICISQODPFGTGAT POSPLMSPRMAHTOSPMWQSQSANPAPSDINGNAQNGNNGNSMFSSQSPHFQGO ANTSMYNNMINVSMATNTGKSSNNQMTGQISMSTSVTSVPTSGLSSMGPEQVNDPA LRGGNLFPNQLPGMDMIKQEGDATRKYC"	
2768..2974	
/gene="TIF2"	
/note="putative alternatively spliced region"	
BASE COUNT 1799 a 1493 c 1406 g 1458 t	
ORIGIN	
misc_feature	
RESULT 9	
HSTIF2GEN	
LOCUS HSTIF2GEN 6156 bp mRNA PRI 10-MAR-1997	
DEFINITION H.sapiens mRNA for transcriptional intermediary factor 2.	
ACCESSION X97674	
NID g1877214	
VERSION . X97674.1 GI:1877214	

Query Match 12.4%; Score 559.4; DB 10; Length 6156;
Best Local Similarity 51.0%; Pred. NO. 1.5e-123;
Matches 2197; Conservative 0; Mismatches 1836; Indels 276; Gaps 26;

QY 64 CAGTTGCTGATGATATCAAGATGAGTATGAGGAGAAA---CTTGATCCACTGGC 120
DB 141 CAGTTGCTGATATGTTTCAAGATGAGTGGATGGGAGAAATACCTCTGACCCCTCCAG 200
QY 121 CAGTGATTCAGAAACGCAAAATGGCAATGTGATCTCCAGGACAAAGGTCTTACCTGCAG 180
DB 201 GGCAGAGACAAAGAAAGCGCAAGGAATGCTCTGACCACTTTGGACCCCAAAAGGAA 260
QY 181 TGGTGAACACGAGAGCGGAGCAGAGAAAGTAATATATATGAGAAATGCTGAGCTGAT 240
DB 261 CACTGAAACAGTAATCGTGAACAGGAAATAAATATATAGAAAGTTCAGAGTTGAT 320
QY 241 ATCTGCCAATCTTGTGATATGACAAATTTCAATGTCTCAACACAGATAAATGTGCGATTT 300
DB 321 TTTTGCATTTTAAATGATATAGCAACTTTTACTTCAACCTGACAAATGTGCAATCTT 380
QY 301 AAAGGAAACAGTAAAGCAGATACGTCAAAATAAAGGACAAAGAAA---AACTATTTCCAA 357
DB 381 AAAGGAACTGTGAAGCAATTCGTGATCAAGCAAGAAAGAAAGCAAGCAGCTGCCAA 440
QY 358 TGATGATGTTTCAAAAGCCGATGTATCTTACAGGCGAGGAGTATTTGATGAAGA 417
DB 441 CATAGATGAAGTGCAGAAAGTCAAGTGTATCTCTACAGGCGAGGTTCTATCGACAGGA 500
QY 418 CTCTTAGGACCGCTTTACTTCAGGCAATGGATGTTTCTTATTTGCTGATGATCGAGA 477
DB 501 TGGCTGGGCGCTATGATGCTTGGGCGCTTGTGTTGTTTCTTCTTGTAGTGAACCTTGA 560
QY 478 GGCAAACATTTGATGATCAGAAATGTACACATACCTGCAATATAGCAAGAGGA 537
DB 561 AGCAACGTTGTGTTCTGTGACAGAAATGTACACATATCTAAGGTATACCAAGAGA 620
QY 538 CTGGTTTAAACAGAGTGTACAAATATCTTACATGAAGAGACAGAAAGATTTCTTAA 597
DB 621 GCTGATGAACAAAGTGTATAGCATCTTGCATGTTGGGGACACACGGAATTTGTCAA 680
QY 598 GAATTTAC---CAAAATCTACATGTTATGAGTTTCTGTCGACAAATGAGCCCAAGACA 654
DB 681 AAACCTGCTCCAAAGTCTATAGTAATGGGGATCTTGTCTGGGCAACCTCCGAGGG 740
QY 655 AAAAAGCCATACATTTAATGCGGTATGTTGATGAACACACCATATCTTCTGGAAGA 714
DB 741 GAACGCCATACCTCAATGTCGGATGCTGGTAAACCTTTACCTGATTCAGAGAGGA 800
QY 715 CATAAAGCCAGTCTGAAATGCGCCAGAGATATGAACAAATGCAAGTGTGCGCTGTC 774
DB 801 GGGTCATGATAACGAGGAAGCTCATCAGAAATATGAACATATGAACTATGACGTCTGCTGTC 860
QY 775 TCAGCCAGAGCTATGATGAGGAAGGGAGATTTGCAATCTGTATGATCTGTGTCG 834
DB 861 TCACCAAAATGCTCATCAAGAAAGAGAGAGATTTGCAAGTCTGCTGATTTGCGTGGC 920
QY 835 ACGCCGATTTACTACAGGAAAGAAACATTTCCATCAACCCCTGAGAGCTTTATACAG 894
DB 921 AAGAGAGTTCCCATGAGGAAGACCACTTCTTCCCTCATCAGAAAGTTTACTACTCG 980
QY 895 ACATGATCTTTAGGAAGGTGTGTAATATAGATACAAATTCATGAGATCCCTCCATGAG 954
DB 981 CCAGGATCTCCAAGCAAGATCAGTCTCTGGATACAGCAACCATGAGAGCGCCATGAA 1040
QY 955 GCCTGGCTTTGAGATATATCCGAAGGTGTATTCAGAGATTTTGTAGTCAATGATGG 1014
DB 1041 ACCAGGTGGAGACCTGTGAAGAGGTGTATTCAGAAATTCATGCGCAGCATGAAG 1100
QY 1015 GCAGTCTATGTTCCCA---GAAACGTCACTATCAAGAAGCTTATCTTAATGCCATGAGA 1071
DB 1101 AGATCTGTCTATGCTTAAAGGCAATCATCATGAAGTACTGAGACAAGGATGGCAT 1160
QY 1072 AACCCCAAGTATATCGATTTCTGTTGGTGTGGAACATATAGTGACTGCACAGACAAAAG 1131

DB 1161 CAGTCAAAATCTATCGTTTTTCCCTGTGATGCACTCTGTGTCGCACAAACGAAGAG 1220
QY 1132 CAAACTCTTCCGAAATCCTGTAAACAATGATGACATGCGCTTTGTCTCAACCCACTTCCT 1191
DB 1221 CAAACTCATCCGTTCTCAGACTACTAATGAACTCAACTTGTATATCTTTACATATGCT 1280
QY 1192 TCAGAGAACAGAGATGGATATAGACCAACCAACCTCCTGTTGGACAAAGGATAGAAC 1251
DB 1281 TCACAGAGACAGAAATGTGTGTGATGAATCGGATCTGACTGGACAAACGATGGGAA 1340
QY 1252 ACCTATG-----GCTGGATGCAACAGTTCGGTAGGGCGCATGATGTGCGCAACCA 1305
DB 1341 GCCACTGAATTCGAATAGCTCTAACAGCCCTGCCCATCAGGCCCTGTGCASTGGGAACC 1400
QY 1306 AGCTTTACAGATCCCGAGCAGCAGAGGCCCTATGGCTTGGCAGACCCCTAGCACCACAGGCA 1365
DB 1401 AGTCAAGACATGACCTCAGTAGCAATATAAATTTTCCCATAAATGGCCCAAGNACA 1460
QY 1366 GATGAGTGGAGCTAGGTATGGGGTTTCCAGTAACATAGCTTTCATTGACCCCTGGGCCAG 1425
DB 1461 AATGGGCATGCCCATG---GGCAGGTTTGTGTCTTGGGGGAATGAACCATGTGTACAG 1517
QY 1426 CATGCAATCAACATCTCTTACAGAACAAACAACTATGGGCTCAACATGAGTAGCCCCCC 1485
DB 1518 CATGCAAGCAACCACTCC---TCAGGGTAGTAATATGCACTCAAAATGAACAGCCCTC 1574
QY 1486 ACATGGAGTCTGCTGCTTGGCCCAACAGCAGAAATATCATATTTCTCTCTGTAATCG 1545
DB 1575 ACAAGCGCCCTGGCATGATCCAGACAGCCCACTCCATGCTTTCCACAGGCATCG 1634
QY 1546 -----TGGGAGTCCAAAGATAGCTCACATCAGTTTTTCTCTGTTC 1587
DB 1635 CATGAGCCCTGGAGTGGCTGGCAGCCCTCGAATCCCAACCCAGTCAGTTTTCCCTGCAGG 1694
QY 1588 AGTGTGCACTCTCCCATGGCATCTTCTGGCAATATCTGGAA---CCACAGCTTTTCCAG 1644
DB 1695 AAGCTTGCATTTCCCTGTGGAGTTTGCAGCAGCAGGAAATAGCCATAGTTATACCAA 1754
QY 1645 CAGCTCTCAGTGGCTGCAAGCCCATCAGTGAAGTGTGGGACATTCCTTTTATCTAC 1704
DB 1755 CAGCTCCTCATGCACTTCAGGCCCTCAGGAGGGGACGGGGTCTCATAGGTCATC 1814
QY 1705 TCTGTCTATCACCAGGCCCAATTTGGATTAATCTCTCC-----AATATGAATAT 1752
DB 1815 GTTGGCTTCACAGACCTCAAAATGGCAATTTGCAAACTCCCACTGTAATATGATCC 1874
QY 1753 TACCAACCAAGTAAAGTAAGCAATCAGGATTCAGAGTCTCTGGGCTTTTATGCGA 1812
DB 1875 TCCCCCACTCAGCAAGATGGAAAGTTGGACTCAAAAGACTGTTTGGACTATATGGGA 1934
QY 1813 CCAAAATCCAGTGGAGTTCATGTGTCAATAGCAGAGATCACTCAGTGACAA 1872
DB 1935 GCCTCTGAAGGTACAACTGGCAAGCAGAGACGCTGCCATCTCTGGAGAGCAAAAGGA 1994
QY 1873 AGAAAGTAAAGGAGC-----AGTGTGAGGGGCGAGAGAAATCAAGGGG 1917
DB 1995 AACAAATGACCCCACTGCCCGCCGCTGAGCAGTGGAGAGCTGACGGGCGAGAGCAG 2054
QY 1918 TCCTTTGGAAGCAAGGTGATATAAAATTTACTGAGTTACTTACTTGTCTTCTGATGA 1977
DB 2055 ACTGCAATGACAGAAAGGGCAGACCAAACTCTCTGAGCTGCTGACCAAAATCTGATCA 2114
QY 1978 CCGGGTCTATCTCTCTGACCAACTCCCCCTAGATTCAAGTTGTAAAGAAATCTTCTGT 2037
DB 2115 GATGGAGCCCTCGCCCTTAGCCAGCTCTTGTGCGGATACAA---ACAAAGACTCCACAG 2171
QY 2038 TAGTGTCAACAGCCCTCTCTGAGTCTCTCTCTATCATCTGAGGAGTATCTCTATCATC 2097
DB 2172 TAGCTGCGCTGG-----TTCTGG 2189
QY 2098 CAATATGATGGGTCACTGTTTACAAGAGAACGACCGGATTTGCAAGTGTGTCAGAA 2157

Dh 2190 GTCTACATGAACCTCGCTCAAGGAGAAGCATAAAAATTTTGCACAGACTCTTGCAGGA 2249
Qy 2158 TGGGAATTCACAGCTGAGGTAGCAAGATTACTGCACAGCCACTGGGAAGACACACAG 2217
Dh 2250 CAGAGTTCCTCCCTGGGACTTGGCAAGTTAACAGAGAAGCCACAGGCAAAAGCTGAG 2309
Qy 2218 -----CAGTAAACTTCTTGTGGGACCGAAATGTTGCAAGCAGGAGCAGCTAAG 2268
Dh 2310 CCAGGAGTCCACGACACAGCTCCTGGATCAAGAGTGAATTAACACAGAGCCGGTGAG 2369
Qy 2269 TCCTAAGAAGAGGAGATAAATGCACTCTTAGATACCTGTGGACAGGATGCTCCAG 2328
Dh 2370 CCCCAGGAAGAAGAG---AATGCACTACTTCGCTATTGCTAGATAAAGATGACTATA 2426
Qy 2329 TGATGCACTCTTAAGAACTACAGCCCAAGTGAAGGAGTGGACATAAATGAGTCA 2388
Dh 2427 AGATATGTTTACAGAAATACCCCAACTTGGAGAGCTGGACAGTAA---GACAGA 2483
Qy 2389 GTGCACAGCTCCAGCATTCCTAGCTCAAGTCAAGAGAAAGACCCCTAAATTAAGACAGA 2448
Dh 2484 TCCGCGAGTACACAAATTAATAGCAATGAAGTGAAGAGGAGAGCTGAGCTTGA 2543
Qy 2449 GACAAGTGAAGAGGATCTGGAGACTTGGATTAATCTAGATGCTATCTTGGTGTCTGAC 2508
Dh 2544 GCCTGGTGACCAAGCCTGGCAGTGAGCTGACAACTTGGAGGAGATTTGGATGATTGCA 2603
Qy 2509 TAGTTCTGACTTTACATAAATTCATATCC-----TCAA 2544
Dh 2604 GAATAGTCAATTACCACAGCTTTCCAGACAGAGGCGGCGCTGTGGATCAGT 2663
Qy 2545 TGGTAGTCACTGGGAGCTTAAGCAACAGGTGTTCAAGGAATTAATCTCTGGGTTGAA 2604
Dh 2664 TGACAAGCAAGCATCATCAATGACCTCATGCACTGAAACAGCCCTGTGAC 2723
Qy 2605 AAGTTCACAGTCTGTGAGTCTATTGCTCCTCCATATAACGAGCAGTGTCTCTGATG 2664
Dh 2724 ACCTGTGGAGCCAGAAACAGCACTGCGAATTTTCACAGAGCACTTTTAATAACCCACG 2783
Qy 2665 CCCTGTTTCTGTGCTCAAGTCTCCAGTAAATAATATCAGTGTCTTCCCATCTTACC 2724
Dh 2784 ACCAGGCAACTGGCAGGTTATTGCCAAACCCAGAAATTTACCACCTTGACATCACATGCA 2843
Qy 2725 AAAGCAACCATGTTGGTGGGAATCCAGAATGATGATAGTACAGGAAATTT----- 2778
Dh 2844 AAGCCCACTGTGTGGACCTTTCCCAACCATCAGAAACATAGTCCCTACTCAGTGAT 2903
Qy 2778 -----ATGGCTCAAGTATGGTGGGCGGCAACCGAAATGTGACTGT 2817
Dh 2904 ACCTCAGCCAGGAATGATGGGTAAATCAAGGGATGATAGGAAACCAAGGAAATTTAGGGA 2963
Qy 2818 GACTCAGACTCTTCTCAGGAGACTGGGCTTACCAACTCAAAAGCGCGCAGAAATGGA 2877
Dh 2964 CAGTAGCAGGAATGATGGTAACAGTCTCTCGGCTACTATGCCATCTGGAGATG 3023
Qy 2878 ACCTATGAATTCAAACTCCATGGGAAGACGAGGAGATTAATATCTCTTTTACCAG 2937
Dh 3024 GGCACCGCAGACTCGGCTGTGAGAGTCACTGTGCTGTACCACAGTGGCATGAACCG 3083
Qy 2938 ACCTGACGTGGTGGCTATTTCACATGTCCTCTTGGTCTAATAGCATACCAGGT--- 2996
Dh 3084 GGCAGTCCAAAGAGGTATGATTGGAACCCAGCAGCCAGCATCCCATGAGGCCAGCAG 3143
Qy 2996 ---GCGACAGCATTTGCAACAGCAGCAGATGCTTCAANTGAGCCGTGGTGAAT 3051
Dh 3144 CCAGCTTGGCCAAAGACAGACGCTTCAGTCTCAGGTTCATGAATATAGGCGCATCTGAAT 3203
Qy 3052 CCCCATGGAATGGGGCTTAATCCCTATGGCCAAAGCAGCAGCATCTAACCAACTGGGTTC 3111
Dh 3204 AGAGATGAACATGGGGACCTCAGTATAGCCCAACAAACAGGCTCTCCAAATCAGACTC 3263
Qy 3112 C---TGGCCCGATGGATTTGTCATGGAACAAGTTTCTCATGGCACTCAAAATAGGCC 3168
Dh 3264 CCGCTGGAAGCATCTCGCTATAGACCAGGCGTCTTTTGGCAGCAACACACAGCATG 3323

Qy 3169 TCTTCTTAGGAATCCCTGGATGATCTTGTGGCCACTTCCRACTTGAAGGCCAGAG 3228
Dh 3324 GCCATTTGGCAGTTCTCGAGATGACTTCTATGTCCACATCTCGCAGTGTAGTCTCCAG 3383
Qy 3229 TGACAAAAGACATTATTGGACAGCTGCACACTCTTCTCAGCAACACAGATGCGACAGG 3288
Dh 3384 TGATGAGGAGCTCTCTGGACAGCTGTATCTGGCCCTTGGGAATTTTGTAT-----GG 3437
Qy 3289 CCTGGAAGAAATGACAGAGCTTTGGGCAATTCCTGAACTTGTCAATCAGGACAGGCAAT 3348
Dh 3438 CCGTGGAGGAGATTGATAGAGCTTAGGAATACCCGAATGCTGACGAGACCAAGCAT 3497
Qy 3349 AGAGCCCAACAGGATGCTTTCCRAAGGCCAAGACACAGCATTAATGATGATCAAGAGC 3408
Dh 3498 AGATCCAGAACAG-----TCTCAAGTCAGATTCCAACTATCATGCTGGAGCAAGAGC 3551
Qy 3409 AGGATTATATGACACATATACCCAGCAGAGGGGCTCCAATGCAAGGAGGCTTTTCATCT 3468
Dh 3552 GCCCGTTTTCCACAGCAGTATGCATCTCAGGCACAAATGCCAGGCTAGCTATTTCTCC 3611
Qy 3469 TCAGGGAACAATCACCATCTTTTAACTCTATGATGAATCAGATGAACAGCAAGGCAAT 3528
Dh 3612 CATGCAAGATCCAACTTTTCAACCATGGAGCAGGGCTAGTTATGCCACACTCCGTAT 3671
Qy 3529 TCCTCTCCAAGGATGCACCCAGCAGCAACATCATGAGACCCCGGACAAACACCCCAA 3588
Dh 3672 CGAGCCAG-----CCGGGCTCAGGCCACGGGCTTAGTGACAGACCCAGCA 3722
Qy 3589 GCAACTTAAATGACATTCAGCAGAGCTGCAGGGCCAGCAGATTTTGAATCAGAGCGG 3648
Dh 3723 TCACTAAGACTTCAACTTCAGCATCGCTCCAAAGCAGCAGCAATCGCCAGC----- 3776
Qy 3649 ACAGCACTTGAATGAAATGGAACCCCTACTCTGGTGGTGTGGTGGTGTAGAGGCC 3708
Dh 3776 -----CACTTATGAATCAATCAGCAATGTTTCCAAATGTGAATTTGACTCTGAG 3824
Qy 3709 TATGATGAGCCCGCAGCAGGTTTCTTAATGCTCAATGCTGCCCAACGACGACAGAGA 3768
Dh 3825 GCCTGGAGTACCAACAGGACCTATTANTGACAGATGCTGGCCCAAGACAGAGGGA 3884
Qy 3769 GCTGCTAAGTCACTCTCCGACACAGAGGCTGCTATGATGATGACAGCAGCAACA 3828
Dh 3885 -----AATCCTGAACCACTCT 3902
Qy 3829 GCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAGCAACAAC 3888
Dh 3903 TCGACAGAGACAAATGTCATCAGCACAGCAAGTTTCAGCAACGAACTTTGATGATGAGG 3962
Qy 3889 CCAGGCTTTCAGCCCACTCTCTAATGTGACTGCTTCCCGCAGCATGATGGCTTTTGGC 3948
Dh 3963 ACAAGGTTGATATGACACCAAGCATGGTGGCTCTAGTGGTATGCCAGCACTATGAG 4022
Qy 3949 AGGACCCCAATGCAAGCTCTCCGCAACAGTTTCCATATCAACCAAAATTTATGGAAT 4008
Dh 4023 CAACCTCGGATTCAGGCAAAATGACACAGTTTCCATTTCTTCCAACTACGGAAT 4082
Qy 4009 GGGACAAACACAGATCCAGCCTTTGGTGGAGTGTCTAGTCTCTCCCAATGCAATGATGTC 4068
Dh 4083 AAGTCAGCAACGTGATCCAGGCTTTACTGGGCTACGACTCCCGCAGAGCCCACTTATGTC 4142
Qy 4059 GTCAAGATGGTCCCTCCAGAAATCCCATGATCAACACCGCAGGCTGCAATCATTA 4128
Dh 4143 ACCCGAATGGCAATACAGAGTCCCATGATGCAACAGTCTCAGGCCAACCCAGCCTA 4202
Qy 4129 TCAG---TCCTCAGAAATGAAGGCTGGCCCTCAGGAAATTTGGCCAGCAAGCTCCT 4185
Dh 4203 TCAGGCCCTCCGACATAAATGGATGGCGAGGGGAACATGGCGGGAACAGCATGTT 4262
Qy 4186 TTCCAGCAGCAGGTTTGGCCACCGAGGGAATCCTGCACTGTATGATATG 4234
Dh 4263 TTCCAGCAGTCCCAACACACTTTTGGCAGCAACACACAGCATG 4311

RESULT 10
MMU39060 4878 bp mRNA ROD 25-APR-1997
LOCUS Mus musculus glucocorticoid receptor interacting protein 1 (GRIPI)
DEFINITION mRNA, complete cds.
ACCESSION U39060
NID G1853979
VERSION 1
KEYWORDS 939060.1 GI:1853979
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1167 to 3560)
Hong H., Kohli K., Trivedi, A., Johnson, D.L. and Stallcup, M.R.
GRIPI, a novel mouse protein that serves as a transcriptional
coactivator in yeast for the hormone binding domains of steroid
receptors
Proc. Natl. Acad. Sci. U.S.A. 93 (10), 4948-4952 (1996)
96209838
2 (bases 1 to 4878)
Hong H., Kohli K., Garabedian, M.J. and Stallcup, M.R.
GRIPI, a transcriptional coactivator for the AF-2 transactivation
domain of steroid, thyroid, retinoid, and vitamin D receptors
Mol. Cell. Biol. 17 (5), 2735-2744 (1997)
97263407
3 (bases 1167 to 3560)
Stallcup, M.R.
Direct Submission
Submitted (20-OCT-1995) Department of Pathology, University of
Southern California, 2011 Zonal Ave. HMR 301, Los Angeles, CA
90033, USA
4 (bases 1 to 4878)
Hong H. and Stallcup, M.R.
Direct Submission
Submitted (12-FEB-1997) Department of Pathology, University of
Southern California, 2011 Zonal Ave. HMR 301, Los Angeles, CA
90033, USA
Sequence and feature updates by submitter
On Feb 28, 1997 this sequence version replaced gi:1314284.
Location/Qualifiers
1. 4878
/organism="Mus musculus"
/strain="ICR outbred"
/db_xref="taxon:10090"
/tissue_type="brain"
/dev_stage="newborn"
204..4592
/gene="GRIPI"
204..4592
/gene="GRIPI"
/function="transcriptional co-activator for steroid
receptors and nuclear receptors"
/note="hormone-dependent interaction with hormone binding
domains of steroid receptors; transactivation activity in
yeast and mammalian cells"
/codon_start=1
/evidence=experimental
/product="glucocorticoid receptor interacting protein 1"
/protein_id="AAC53151.1"
/db_xref="PID:g1853980"
/db_xref="GI:1853980"
/translation="MSGMGENTSDPSAETRRKKECDQLGSPKSTKRRNEQENK
YIELADLIPANFNDINFNFKDKCAILKTYKQIRQKEQKAAANIDEVQKSDV
SSTGGVIDLALGPMLEALDGFVFNLEGSVFSVNTVYLYRNOELMKNKSY
SILHVGDTFVNLKPKSMVNGSGSSEPRSTHTFNCRLMKVLPDSEEGHDSQ
EAHQYEAMCFVSPQKSIKEGEDQLQCLICVARRVPMKERTLPSSSEFTTODL
QKGTISLDTMRAAMKPGWEDLVRRICQKFTQHEGESLYAKRHHHEVLRGLAFS
QIVRFSISDGLVAQTKSLIJSQTNELPOLVLSLHMLHREQNVCMNPDLTGQAMG
KPLNPISSSPAHQALCSGNPGQDMILGSINFPNMGPKQMGKRGFGSGGMHV
SGMQATPOGNSYALMSPSQSPGMNPGQASVLSIPRMSPGVAGSPRIPPSQFS
PAGLSHPVGVCSSSTGNSHTYSSLSNALQALSEGHSVLSGLSLASPDLMKGNLQNSP

VNNPPPLSKMSLDSKDCFLYGESEGTGQAEASCHPEEKQNDSSMPQASGD
RAEGHSLRHSKQTKLLQLATKSDMEPSPLSLSDTKNDKSLGSPGSGTHGS
LKEKHRLHLLQDSSPDVLAETAEATKELSDQSSSTAPGSEVTVKQEPASPKKK
ENALLRYLLDKDDTKDIGLPETPKLERLDSKTDPSANTKLIAMKTVKEVSEPSDQ
PGSELNLEIILDDLQNSOLPOLFPDTPRCAPTGSVDKQAIINDLMQLTADSSPVA
GAQKALRMSQSFENNPRGQLGRLLPNQNLPLDITLQSGTGAFFPIRNSPFSYVI
PQGMNGNOMLGSQGNLGNSTSGSSTSRPSPGSPWAPQSPVRVTCATATGAM
NRPVQGMIRNPITASIPMRANSOPQRMQSOVMNIGSELMNMGPOYNQOQAPP
NOTAPWESILPIDQAFASQNRPFSSPDLLCPLPAAESPDGALIDQLYLALR
NFDGLIEDRALGIPELVSQSAVDAEAFSSOESSIMLEKPKPVFPQOYASOAOAQQ
GYNPMODPNFTMGORPNTTLMOPRGLRPTGIVQNOPNOLRLOHRLQAOQNRQ
PLMNQISSYNNVNLTLRPGVTPQAPINAOHQAQRREILNHLRQMQOQVQORILM
MRQGLNVTSPSWAPAGLPAASNPRIQANAQOFPFPNPNYISQOPDPEGTGATTPQ
SPLMSPMAHTOSPMQOANPAYOTSDMNGWAQSGMGSFMFSQOQSPHPFGQOAN
TMSYNNMNISSVMATITGLSSNMONTQMSMTSVTSVPTSLPGLSPNGPQVNDPALR
GGNLPNQLPGMDMIKQEGDASRKYC"
BASE COUNT 1312 a 1381 c 1256 g 926 t 3 others
ORIGIN
Query Match 11.5%; Score 517; DB 12; Length 4878;
Best Local Similarity 51.1%; Pred. No. 2e-113;
Matches 2204; Conservative 0; Mismatches 1815; Indels 298; Gaps 30;
QY 64 CAGTTGCTGATGATATTCAGATGAGTGGATTAGGAGAAAACCTTGGATCCACTGGCCAG 123
DB 182 CAGTTGCTGATGATGTTCAAGATGAGTGGGATGGGAGAAAACACCTCTGACCCGCTCAG 241
QY 124 T---GATTACAGRAAAGCGCAATTCATGATCTATCCAGGACAAGGTCTTACTCTGAG 180
DB 242 GGCAGACAGCAAGAAACGCAAGAAATGTCGCCAGCACCTCGGACCCAGCCCAAGAGAG 301
QY 181 TGGTGAAGAAACGGAGACGGAGGAGAAAGTAAATATATTGAAGAAATGGCTGAGCTGAT 240
DB 302 CACTGAGAAACGAGACCGCGAGGAGAGATAGTACATAGAGGAGCTGCCGATCTGAT 361
QY 241 ATGTGCAATCTTCTAGTATGATTTGACAAATTTCAATGTCTCAACACAGATAAATGTGCGATTTT 300
DB 362 CTTCGCAAACTTTAATGATATTGACAACTTCAACTTCAACCTGACAAATGTGCCATCT 421
QY 301 AAGGAAACAGTAGACAGATACGTCAATTAAGAGACAGAGAAA---AACTATTTTCAA 357
DB 422 AAAGAAAACTGTGAAGCAGATCCGCCAGATCAAGAGACAGAGAAAGACAGCTGCCAA 481
QY 358 TGATGATGATGTTCAAAAACCGCATGTATCTCTACAGGCGCAGGAGTATTATGATAAGA 417
DB 482 CATAGATGAAGTCAGAAAGTCAGATGTCTCCACGGGCGAGGCTGTCATCGACAGGA 541
QY 418 CTCTTAGGACCGCTTTACTTCAGGCATTTGGATGGTTTCTTATTTTGTGTGAATCGAGA 477
DB 542 TGCACCTGGGCCCATGATGCTTGAGGCCCTCGATGGGTTCTTCTTCTGTTGTAACCTGGA 601
QY 478 GGCNAACATTTGATTTGATATCAGAAATGTCACAAATACCTGCAATATAGCAAGAGGA 537
DB 602 AGGCAGTGTGGTGTCTGTGTCAGAGAAATGTGACACAGTATCTACGGTATATACCAAGAAGA 661
QY 538 CCTGTTTAACACAAGTCTTTACAAATATCTTACATGAAGAGACAGAAAGATTTTCTTAA 597
DB 662 GCTGATGAACAGAGTGTCTACAGCATCTGTCATGTCGGGACCCACACTGAATTTGTCAA 721
QY 598 GAA---TTTACCAAAATCTACAGTTAATGGAGTTTCTCGGACAAATAGAGCCCAAGAGACA 654
DB 722 GAACCTGCTGCCAAGTCCATGTTGAATGATGAGGATCCTGGTGGAGAACCTCCAGGCG 781
QY 655 AAAAAGCCATACATTTAATTTGCCGTATGTTGATGAAGAACACCATGATATTTCTGGAGA 714
DB 782 GACGACCATACCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
QY 715 CATAAAGCCAGCTCTCAATGCGCCAGAGATATGAACATGACGTGCTTGGCCCTGTC 774
DB 842 AGGCCATGATAGCCAGGAGAACCCATCAGAAATACGAGGCGATGCGATGCTGCTGCTGCTGCT 901
QY 775 TCAGCCACGAGCTATGATGAGGAAGGGAAGATTTTGAATCTTGTATGATCTGCTGCTGCTGCT 834

D	b		902	TCAGCCCAAGTCCATCAAAGAGAAAGGCGAAGATTGCAGTCTCGTTGATTGTGTGGC	961
Q	y		835	ACGCCGATTACTACAGGAAGAAGAACATTTCCATCAAAACCCTCAGAGCTTTATTACCAG	894
D	b		962	ACGAAGTGCCCCATGAAGAAAGACCAACTTCCCTCATCAGAAAGCTTTACCACCG	1021
Q	y		895	ACATGATCTTTCAGGAAGGTTGTCAATATAGATAACAATTCAGTAGATCTCTCCATGAG	954
D	b		1022	CCAGGACCTCCAAGGCAAGATCACCTCAGTGACACTAGCACCATGAGAGCGCGCATGAA	1081
Q	y		955	GCTTGGCTTTGAAGATATAATCCGAAGGTGATTTCAGAGATTTTTTAGTCTTAATGATGG	1014
D	b		1082	GCCGGCTGGGAAGATCTGGTAAAGAATGCAATTCAGAAGTTCCACACACAGCATGAAG	1141
Q	y		1015	GCAGTC---ATGCTGCCAGAAACGTCATATCAAGAAAGCTTATCTTAATGSCCATGCAGA	1071
D	b		1142	GGAGTCTCTATCATATGCCAAGAGGCATCACCATGAAGTCTGAGACAAGGTTGGCGTT	1201
			1072	AACCCAGATATATCGATTCTCGTTGGCTGATGGAACTATAGTGACTGCACAGACAAAAG	1131
			1202	CAGTCAGATCTATCGTTTTTCTTGTCTGATGGCACTCTCGTCTGCACAAACCCAGAG	1261
Q	y		1132	CAAACTCTCCGAAATCTGTAAACAAATGATCGACATGGCTTTGTCTCAACCCACTTCCT	1191
D	b		1262	CAAACTCATCCGTTCTCAGACTACTAATGAGCCTCAGCTTGTAAATATCTTTACACATGCT	1321
Q	y		1192	TCAGAGAGAACAGATGGATATAGACCAAAACCAAACTCTCTTGGACANAGGA-----	1245
D	b		1322	TCAGAGAGCAGAAATGTATGTATGAATCCGGATCTGACTGGACAGGGATGGGGAA	1381
Q	y		1245	-----TTAGACCACCTATGGCTGGATGCAACAGTTTCGGTAGCGGCATGAGTATGTCGC	1299
D	b		1382	GCCATTGAATCCAATTAGTCTTAGCAGCCCTGCCACCAGGCCCTGTGCAGTGGGAACC	1441
Q	y		1300	AAACCAAGGCTTACAGATCGCAGCAGCAGGCGCTATGGCTTGGCAGACCCCTAGCACCC	1359
D	b		1442	AGGTCAGGACATGACCCCTCGTAGCAATATAAATTTCCCATGAATGCCCCAAGGAACA	1501
Q	y		1360	AGGCAGATGAGTGAGCTAGTATGGGGTTCCAGTAACATAGCTTCATGACCCCTCG	1419
D	b		1502	AATGGGCATGCCTATGGCAGAGTTTGGTGGTTCTGGGGGCA-----TGAACCATGT	1552
Q	y		1420	GCCAGGATGCAATACCACTTCTTACAGAAACAACATATGGCTCAACATGAGTAG	1479
D	b		1553	GT CAGGCATGCAGGCAACCACTCC---TCAGGTTAGTAACATATGCATCAAAATGAACAG	1609
			1480	CCCCCCACATGGGAGPCTTGCTTGCCCCAAACACAGACAATATCATGATTTCTCTCTCG	1539
			1610	TCCTCGAAAGCAGCCCGGCATGAACCGGGGCAAGCCAGCTCCGTGCTCTCCCCAAG	1669
Q	y		1540	TAAATCG-----TGGAGTCCAAGATAGCTTCGGCAATACTCGGAA--CCACAGCTT	1638
D	b		1670	GCAGCCATGAGCCCGCGGTGGCTGGCAGTCTCGCATCCACCCACATGATTTTCCCC	1729
Q	y		1582	TGTTGAGGTGTGCACTCTCCCATGSCATCTTCTGGCAATACTCGGAA--CCACAGCTT	1638
D	b		1730	TG CAGGAGCTTGCAATCCCTGTGGGAGTTTG CAGCAGCAGACAGAAATAGCCATAGTTA	1789
Q	y		1639	TTCCAGCAGCTCTCTCAGTGCCTCGAACCCATCAGTGAAGGTTGGGGAGCTTCCCTTTT	1698
D	b		1790	TACCAACAGTTCCCTCAATGCACTGCAAGCCCTCAGCAGGCGCATGGGTCTCACTCGG	1849
Q	y		1699	ATCTACTCTGTATACACAGGCCCAATATGGATNACTCTCCC-----AATAT	1746
D	b		1850	GTCTCTGGCTGTTCACCGGAGCTAAAAATGGCAATTTGCAAAACTCCCCAGTTAAT	1909
Q	y		1747	GAATATTACCCRAACAGTAAGTAGGCAATCAGGATTCCAAGAGTCTCTGGGCTTTTA	1806
D	b		1910	GAATCTCCCCCACTCAGCAGAGTGGGAAGCTTGGACTCCAAAGACTGTTTGGACTTTTA	1869
Q	y		1807	TTGCGACCAAAATPCCAGTGGAGGATTCAATGTGTGAGTCAAAATAGC-----	1853

QY 2871 -GAATGGACCTATGATTCACAACTCCATGGGAAGACACGAGGAGATTAATTAATCTCT 2929
DB 3059 GGAATGGGACACACAGAGTCCAGCT--GTGAGAGTCACTTGTGCTGCTACCACTGGTGCC 3116
QY 2930 TTAACACAGACCTGCACTGGGTGGCTATT-----CCACATGCTCTTGGTCTTAATA 2984
DB 3117 ATGACCCAGCACTCAAGGAGGATGATTCGGAACCCACAGGAGGATCCCATGGGA 3176
QY 2985 GCATACAGGTCGAGAC--AGTATTGCAACAGCAGCAGAGATGCTTCAATGAGGCT 3043
DB 3177 GCAACAGCAGCCTGGCCAAAGACAGATGCTTCACTGCTCAGTCAATGAACATAGGCCCT 3236
QY 3044 GGTGAATCCCCATGGGAATGGGGCTATCCCTATGGCCAA---GCAGCAGCATCTAAC 3100
DB 3237 TCTGAGTTAGAGATGAACATGGGAGGACCTCAGTATAATCAACAGCAGGCCCCCTCCGAC 3296
QY 3101 CAACCTGGGTTCTGCGCCGATGGCATGTTGTCATGGGAACAGTTTCTCATGCACTCAA 3160
DB 3297 CAACCTGCCCCGTGGCTGAGAGCATCTGCTATAGACCGGATCGTTTCCCGCCGAG 3356
QY 3161 AATAGGCTCTTCTTAGGAATTCCTGGATGATCTTGTGGCCACCTTCCAACTGGAA 3220
DB 3357 AACAGGAGCCTTCGGCAGCTCCCTGATGACCTGCTGTGTCACATCTCCAGCAGAG 3416
QY 3221 GCCAGAGTGAGCAAGAGCATATTGGACAGCTGCACACTCTTCTCAGCAACACAGAT 3280
DB 3417 TCGCCAAAGCGATGAGGGGCTCTTCTTGACCACTGTATCTGCGCTTGGGAACCTTCGAT 3476
QY 3281 GCGACAGGCTGGAAGAATTCACAGAGCTTGGCACTTCTGCACTTGTCAATCAGGA 3340
DB 3477 -----GGCTTGGAGGATTCATAGAGCTCTGGGATACCAAGACTGTGTCCAGCAGC 3530
QY 3341 CAGGCATTAGAGCCCAACAGATGCTTTCCAGGCCAAGAGCAGCAGTATGATGAT 3400
DB 3531 CAAGCTGTGATGAGCAGAG-----TTCTCAAGTCAGGATCCAGATATGCTGAG 3584
QY 3401 CAGAGGAGGATTTATGAGCAGACATACCCAGCAGAGGGGCTTCCAAATGCAAGAGGC 3460
DB 3585 CAGAAGCCCCGTTTCCACAGCAGTACGCATCTCAGCAGCAAAATGGCCCCAGGTTGC 3644
QY 3461 TTTCACTTCAGGCAATCACTCTTTTAACTCTATGATGAATCAGATGACCAAGCA 3520
DB 3645 TATAATCCCATG---CAAGATCCAACTTTTACACATGGGACAGCGCCAAATACACC 3701
QY 3521 GGCATTTTCTCTCCAGGAATGACCCACAGCAGCAACATCATGAGACCCCGGACAAAC 3580
DB 3702 ACACCTCCGTATGACGCGCCGCGGCTCAGGCCACAGGATGTAC-----AGAAC 3755
QY 3581 ACCCCCAAGCACTTAGAATGAGCTTCAGCAGAGGCTGCGGCGCAGCAGTGTGAT 3640
DB 3756 CAGCCAAACCAACTGAGACTTCAGCTTCAGCACCGCTCCAGCAGCAGAGAAATGCCAG 3815
QY 3641 CAGAGCCGACGCACTGAATGAAATGGAATGGAACCCCTACTGCTGGTGGTCTCGGTG 3700
DB 3816 CCG-----CTTATGATCAGATCAGCAGTGTTCCTCAATGTGAACCTG 3857
QY 3701 ATGAGGCTATGATGAGCGCCAGCAGGCTTTCTTAATGCTCAATGCTGCGCCAAAGC 3760
DB 3858 ACTCTGAGGCTTGGAGTGGCCACTCAGGCTCTATTAATGACAGATGCTGCGCCAGAGG 3917
QY 3761 AGCAGAGAGCTCTAAGTATCATCTTCGACAAACAGAGAGGCTGCTATGATGACGAGC 3820
DB 3918 CAGAGGAAATCCTCAACCAACCTTCG----- 3947
QY 3821 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3880
DB 3947 -----GCAGAGGAGAGATGACAGCAGAGGTCAGCAGCGGACTCTGATG 3989
QY 3881 CAGCAAAACCCAGGCTTACAGCCCACTCTTAATGTGACTGCTTCCCGCCAGCATGGG 3940
DB 3990 ATGAGAGGAGCAGGCTTGAATGTGACCCAAAGCATGGTGGCTCCCGCTGCGCTACAGCA 4049

QY 3941 CTTTGGGAGGACCCACAAATGCCAACAGCTCTCCGCAACAGTTTCCATATATCAACCAAT 4000
DB 4050 GCATGAGCAATCCCGGATCCCGAGCCAAATGCCAGCAGTTTCCATTTCTCCGAC 4109
QY 4001 TATGGAATGGGACAAACACAGATCCAGCTTTTGGTCCAGTGTCTAGTCTCCCAATGCA 4060
DB 4110 TACGGAATAGTCAACACCTGATCTGCTTTACTGGGTACACCTCCCGCAGAGTCT 4169
QY 4061 ATGATGCTCAAGAAATGGTCCCTCCAGAAATCCCATGATGCAACACCGCAGCTGCA 4120
DB 4170 CTAATGCTCCCGGATGTCACATATCTCAGAGTCCCATGATGACAGCTCTCAAGCCAC 4229
QY 4121 TCCATCTATCAG---TCTCAGAAATGAAGGCTGCGCATCAGGAAATTTGGCCAGAAC 4177
DB 4230 CCAGCTACCAAGCCACCTCAGCATGAATGATGGACAGGAGGATGCTGGTGAAC 4289
QY 4178 AGCTCTTTTCCAGCAGCAGTTTCCGACAGGGAATCTCGCAGTGTATAGTATG 4234
DB 4290 AGCATGTTCTCACAGACATCCCAACACATTTGGGCAACAGCAACACCAAGCATG 4346

RESULT 11

AF136943
LOCUS AF136943 4398 bp mRNA ROD 12-APR-1999
DEFINITION Rattus norvegicus transcriptional intermediary factor 2 (TIF2)
mRNA, complete cds.
ACCESSION AF136943
NID 94581052
VERSION AF136943.1 GI:4581052
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
REFERENCE 1 (bases 1 to 4398)
AUTHORS Leirs,J., Treuter,E. and Gustafsson,J.-A.
TITLE Mechanistic principles in NR box-dependent interaction between
nuclear hormone receptors and the coactivator TIF2
Mol. Cell. Biol. 18 (10), 6001-6013 (1998)
JOURNAL 98414616
MEDLINE 2 (bases 1 to 4398)
AUTHORS Leirs,J., Treuter,E. and Gustafsson,J.-A.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1999) Department of Biosciences, Karolinska
Institute, Haelsovaegen 7, Huddinge 14157, Sweden
FEATURES Location/Qualifiers
source 1..4398
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
gene 1..4398
/gene="TIF2"
CDS 1..4398
/gene="TIF2"
/function="nuclear hormone receptor coactivator"
/codon_start=1
/product="transcriptional intermediary factor 2"
/protein_id="AAD24587.1"
/db_xref="PID:94581053"
/db_xref="GI:4581053"
/translation="MSGMGNTSDPSRAETRRKECPDQLGPSKRSTKRNROENK
YIEELALIFANFDNFNFKDCAILKQIRIQKEKAAANIDEVQKSDV
SSTGGVIDKALGPMLLEADGFFVNLGNVVFVSENVQVLRVNOELMKNKSV
SILHVGDHTEFVKLLPKSMVNGSWTGEPRRNSHTFNCRLMKPLPDSSEEGHDND
EAGQYETMQCFAYSPQKSIKEGEDKQSLICVARVPKMERALPSSEFTTRQDL
QGKTFIDTSMRDAMPWEDLVRCIKQFHTQHEGESLSARKHHHEVLRQGLAFS
QIYRFLSDGTLVAATKSLIRSQTTNEPOLVISIHLHREQVNCVNMNPDLTQAMG
KPLSPSSSSSPARQACGNPQDVALGNSMNFPMNGPREQMSPMRFGSGGMNHV
SGMAHTTPQGSNVALKWNPSQSLNPGQSSVLSPRHRMSPGVRSPVPSQFS
PAGSLHSPAGVCSSTGNSHSYTNLSNALQALSEGHVSLGPSLASPLKMNQNSP
VNMNPPLSKMSLDKDCFLYGEPEGCTGQAQASCHPEEQKRPNDSSMPQASD
RAGHSRLHESKGTQKLLQLLTTKSDQMEPSLPSSLSDTNKDSTGSLPGSGTHGTS
LKEKHKLHLRLQDSSFDLAKLTAEATKELNSESSTGAPSEVTVKQBPSPKPK
ENALLRYLLDKDDTKDGLPEITPKLERLDSKTDPSNKLKIANKTVEVSEFSPDQ

PGSELDNLEIBLDDLQNSQLPQLPDPTRPGAPTGSVDKQAIINDLMQLTADSSPVTVP
GAQKAAIRMGSQTFNNRPGQLGRLLPNQLDLTLOSPTGAGPPPIIRNSPYSVI
PQPMQNMQGLSGQNLGNSTMGTSRSRSMFSGEWAQSPVRVTCATTGAM
NRTIQGMIENRTPIMRANSQGRQMLQPMVNI GPSELMNMGGPQYNOQAAPP
NOTAPNESTLPIDQASFGSONRHPFGSPDDLLCPHPAAESPDECALDQYLALR
NFDGLIDRALGIPELVQSQADPEQFSQSSNMLEOKPVFQVQASOTQMAQG
SYNPMDDPNFTWQGRNPNTLBMQPPRGLRPTGIYONPNOLRLQLOHLQAQNRQ
PLNQISGVSNVNLTRPGVPTQAPINQAMLAQRREILNQHLRQKHQHQVQOQRT
LMRGGLGNLTSMVAPTLGLPAMNSNPRIPQANAQFPFPFPNYGISQOPDPGTGAT
PQSLPLMSPRMAHTQSPQMWOQSANPAQPSADINGWAQSGMGNFSQSOPHPFGQ
ANTSMYNNNNINIAGLSNNNNQNTQMSMTSVTSVPTSGLSMSMGEQVNDPA
LRGSSLTFTNQLPGMDMIKQEGDGSRRKYC"

1233 a 1232 c 1125 g 808 t

BASE COUNT
ORIGIN

Db	781	AGGCCAGCCCTTCCCTCATCAGAGAGCTTTACACCCGCCAGGACCTCTCAAGGCAAGATC	840
Qy	917	GTCAATATAGATACAAATTAACCTGAGATCCTCCATGAGGCGTGGCTTTGAAGATATATATC	976
Db	841	ACTTTCTGGACACTAGCACATGAGAGATGCCATGAACCGGCTGGGAGGATCTGGTA	900
Qy	977	CGAAGGTGTAATCAGAGATTTTTTAGCTTAAATGATGGGCAGTC---ATGTCCTCCAGAAA	1033
Db	901	AGAGATGTATTTCAGAAGTTCACACACAGCATGAAGGGAGTCTCTATCATACGCCAAG	960
Qy	1034	CGTCACATCAAGAAGCTTATCTTAATGGCCATGAGAAACCCCAAGTATATCGATTCCTCG	1093
Db	961	AGGCATCACCATGAAGTTCGTGACAAAGGCTTGGCGTTCAGTCAGATCATCTGTTTTTCC	1020
Qy	1094	TTGGCTCATGGAACTATGTGACTGCACAGACAAAACCAACTCTTCCGAAATCCTGTA	1153
Db	1021	TTGTACAGGGCACTCTCGTTGTGCACAAAGCAAGCAACTCATCCGTTCCAGACT	1080
Qy	1154	ACAAATGATCGCATGGCTTGTCTCAACCCACTTCTTCAGAGAGAACAGAATGGATAT	1213
Db	1081	ACGAATGAGCCTCAGCTTGTGATATCGATACACATGCTTCACAGAGACAGAATGTGT	1140
Qy	1214	AGACAAACCCAAATCCTGTGTGACAAGGATTT-----AGACCACCTATGGCT	1261
Db	1141	GTAATGAATCCGATCTGACTGGCAAGCAATGGGAAGCCATTGAGTCCAATGAGTCT	1200
Qy	1262	GGATGCAACAGTTTCGTAGCGGCGATGAGTATGTGCGCAACCAACCAAGGCTTACAGATCGCG	1321
Db	1201	AGCAGCCCGCCCGTCAGGCGCATGTGCACTGGGAACCCAGTCAGGAGCTGGCCCTCGGT	1260
Qy	1322	AGCAGAGGCGCTATGGCTTGGCAGACCTAGCACACACAGGCGCAGATGAGTGAGCTAGG	1381
Db	1261	AGCAATATGAACHTTCCCATGAATGGCCCAAGGGAACAGATGAGCATGCCCTATGGGAGG	1320
Qy	1382	TATGGGGTTTCCAGTAACATAGCTTCANTGACCCTTGGGCCAGGCATGCAATACCACTCT	1441
Db	1321	TTTGGTGGTCTGGGGCA-----TGAACATGTATCAGGCATGCAAGCAACCACT	1371
Qy	1442	TCCTACCAGAACAACTATGGGCTCACTAGTAGTACGCCCCCNCATGGGAGTCTCTGGT	1501
Db	1372	CC---TCAGGGTAGTAACATGACATCAAAATGAACATGCTCCTCACAAGACGCCCCGGC	1428
Qy	1502	CTTGCCCAACACAGAGAAATCATGATTTCTCTCTGTAATCG-----	1546
Db	1429	CTGAACCCAGGGCAACCCAGCTCTGTCTCTCCCAACGCATCGCATGAGCCCCGGGCTG	1488
Qy	1546	--TGGAGTCCAAGATAGCCTACATCAGTTTTTCTCTGTGTCAGGTGTGCACCTCTCCC	1603
Db	1489	GCTGGCAGTCTCCGCTTCACCCAGTCAGTTTTCCCTGTCAGGAAGCTTGCATCTCCCA	1548
Qy	1604	ATGCGCATCTCTGGCAATACTGGGAA---CCACAGCTTTTCCAGCAGCTCTCTCATGTGCC	1660
Db	1549	CGGGAGTTTTGCAGCAGCAGAGAAATAGCCATAGTTACACCAACAGTTCCTCTCAATGCA	1608
Qy	1661	CTGCAAGCCATCAGTGAAGGTGTGGGACTTCCCTTTTATCTACTCTGTCATCACCAGGC	1720
Db	1609	CTGCAACCCCTCAGCAGGGGCCAGGGGTCTCGTTAGGGCCATCGTTGGCTTCTCCGGAC	1668
Qy	1721	CCCAATTTGGATAACTCTCCCAAT-----ATGAATATTACCAACCAAGTAAA	1768
Db	1669	CTAAAATGGCAATTTCGCAAAACTCCCCAGTTAACTGAATCTCTCCCCCACTCAGCAAG	1728
Qy	1769	GTAAGCAATCAGGATTCGAAGAGTCTCTGGGCTTTTATTCGGACCAAAATTCAGTGGAG	1828
Db	1729	ATGGGAAGCTTGGACTCCAAAGACTGTTTGGCACTTATGGGGAGCCATCAGAAGGTACA	1788
Qy	1829	AGTTCAATGTCTCAGTCAAAATAGCAG-----AGATCACCCTCAGTGACAAAAGTA--	1881
Db	1789	ACTGGCAACGACAGGCCAGCTGCCATCCTGAAGAGCAGAGAGGCCCAATGATTCNCAG	1848
Qy	1881	-----AGGAGACAGTGTGTAGGGGCGCAGAGAAATCAAAAGGGTCTCTTGGAAAGCAA	1933

Db 1849 ATGCCCCAGCGCGCCAGCAGGACAGGGCTGAGGACACAGCGGCTGCATGAGACAAA 1908
QY 1934 GGTCTATAAAAAATTAATGTCAGTACTTACCTGCTGTTCTTCTGATGACCGGGTCAATCTCTCC 1993
Db 1909 GGGCAAAACCAACTCTCTACAGCTGCTCACCACCAAGTCCGACAGATGGAGCCTTCACCC 1968
QY 1994 TTGACCAACTCCCCCTTAGATTCAAGTTGTAAGTAATCTTCTGTAGTGTACACAGCCCC 2053
Db 1969 TTGCCCAAGCTCTTTGTCGGACAAA----- 1994
QY 2054 TCTGGAGTCTCTCTCTACATCTGGAGGAGTATCTCTACATCCAAATATGATGGGTCA 2113
Db 1994 -----ACAAGGACTCCACAGGAGCTTCCAGGGGCTGGGTCCAGCATGGCACC 2043
QY 2114 CTGTTTACAAGAGAACCGGATTTTGCACAAGTTGCTGCGAGAAATGGGAATCACCAGCT 2173
Db 2044 TCCTCAAGGAGAGCAATAAGATTTTGCACAGACTCTTTACAGACAGCAGCTCCCTGTG 2103
2174 GAGGTAGCCCAAGATTACTGACACAGCCACTGGGAAGACACCA-----GCAGTATA 2224
Db 2104 GACTTGGCCAAAGCTGACAGCAGAGGACACAGGCAAGAGCTGAATCAGGAGTCCAGTGGC 2163
QY 2225 ACTTCTTGTGGGACGAAATGTTGTCAAGCAGGAGCAGCTAAGTCTTAAGAAAGAGGAG 2284
Db 2164 ACGCTCTCTGGTTCAGAACTGCTGTCAACAGAGGCCAGCGGACCCAGAAAGAG 2223
QY 2285 AATAATGCACTCTTTAGATACCTTGTGGACAGGAGTATCTTGTAGTATGACACTCTCTAAA 2344
Db 2224 AAC---GCACCTACTGGGCTATTTGTCGACAAAGATGATACTAAAGATATTGTTTACCG 2280
QY 2345 GAACCTACAGCCCCAAGTGGAGAGTGGCAATAAATGATGATGTCAGCCACTCCACC 2404
Db 2281 GAATAACCCCAAACTCGAGCGGGTGGACAGTAA---GACAGATCTTGGCCAGTAACACA 2337
QY 2405 ATTCTTAGCTCAAGTCAAGAGAAAGACCCCTAAATTAAGACAGACAGCAAGTGAAGGGA 2464
Db 2338 AAGCTAATCCCATGAAACTGTGAAGGAGAGGTGAGCTTTGAGCCAGTGACAGCCT 2397
QY 2465 TCTGGAGACTTGGATTAATCTAGATGCTATTCTTGGTGAAGTCTGACTAGTCTGACTTTAC 2524
Db 2398 GGCAGCGAGCTGACAACTTGGAGAGATTTTGGATGATTGTCAGAAATAGTCAGTTACCA 2457
QY 2525 AATAATTCATATCCCAATAGTGTAGTCATC-----TGGGGACTAAGCAAGGTGTTTC 2579
Db 2458 CAGCTTTTCCAGACACAGGCGCAGGAGCTCTACTGGGTGAGTTGACRAAGCAAGCCATC 2517
2580 AAGGAACATAATCTCTGGGTTTGAAGATTACAGTCTGTGCAAGTCTATTTCGTCCTCCAT 2639
Db 2518 ATCAATGACCTCATGCACTACGGCTGACAGCAGTCTGTACACTGTCTGGAGGCCAA 2577
QY 2640 ATAACCGAGCAGTGTCTCTGGATAGCCCTGTTT-----CTGTTGGCTCAAGTCCCTC 2690
Db 2578 AAAGCAGCACTGCGAATGTACAGAGACACTTTTAAATAACCCAGCACCAGGCAACTGGGC 2637
QY 2691 CAGTAAAAATATCAGTCTTTCCCATGTATACCAAGCAACCCATGTTGGGTGGGAATC 2750
Db 2638 AGGTATTGCAAAACCAAGAAATTTACCACTTTGACATCAGATTCGAAAGCCCAACTGTGCT 2697
QY 2751 CAAGATGATGGATAGTACAGAAAAATTTATGGCTCAAGTATGGTGGGCCAAACCGAAATG 2810
Db 2698 GGACCTTTCCCAACCAATCAGAAACAGTAGTCCATACTCAGTGATACCTCAGCCAGGAA-- 2756
QY 2811 TGACTGTGACTCAGACTCTCTCTCAGGAGACTGGGGCTTTACCAAACTCAAAGCGCGGCA 2870
Db 2756 TGATGGGTACCAAGGAGTGTAGGAAGCCAAAGAACTTTAGGNAACATAGCACAGGAA 2815
QY 2871 GAATGGAACCTATGAATTCAAACTCCA----- 2898
Db 2816 TGATGGCAGTAGCACTTCCCGGTCAGCATGCTTCTGGGGAATGGGCACACAGATC 2875
QY 2898 -----TGGGAAGACCGAGGAGATTAATACTCTTTTACCAGACCTGCACTGGGTG 2951
Db 2876 CAGCTGTGAGAGTCACTTGTGCTGCCACCACTGGTGCCATGAACCGGCCAATCCAAGGGG 2935

QY 2952 GCTCTAAT-----CCACATTTGCCCTCTTTGGTCTTAATAGCATATACCAGTGGGAGACCAG 3005
Db 2936 GCATGATTCGGAAACCCCAACAGCAAGCATCCCATGGAGCAACACAGCAGCTTGGCCAAA 2995
QY 3006 TATTGCAACAGCAGCAGCAGATGCTTCAATGAGGGCTTGGTAAATCCCATGGGAATGG 3065
Db 2996 GACAGATGCTGACGCGCTCAGGTCAATGAACATAGGGCCTTCTGAGTTGGAGATGAACATGG 3055
QY 3066 GGCTAATCCCTATGGCCAA---GCAGCAGCATCTCAACACTGGGTCTCTGCCCCGATG 3122
Db 3056 GAGGCCCTCAGTATATCAACAGCAGGCCCCCTCCAAACCAAACTGCCCCCTGCCCCGAGA 3115
QY 3123 GCATGTTGTCATGGAACAAGTTTCTATGGCACTCAAAATAGGCTCTCTCTTTAGGAAT 3182
Db 3116 GCATCTGCTATAGACAGGATCTTTGGGACGCCAGAACAGGCATCTCTTTGGCAGCT 3175
QY 3183 CCTGATGATCTTGTGGGCCACCTTCCAACTGGAAGGCCAGAGTGAAGAAAGAGCAT 3242
Db 3176 CCCCCAGCAGCTGCTCTGTCCACATCTTGCAGCCGAGTCCCCAANGCAGGAGGGCGCTC 3235
QY 3243 TATTGACCCAGCTGTCACACTCTTCTCAGCAACACAGATGCGACAGGCTTGGAAAGAAATG 3302
Db 3236 TTCTAGACCACTCTATCTGGCTTTCGCGAATCTCGAT-----GSCCTTGAAGAGATTG 3289
QY 3303 ACAGAGCTTTGGCATTCTCTGAATTTGTCAATCAGGAGCAGGCAATTAGAGCCCAACAGG 3362
Db 3290 ATAGAGCTCTGGGATACCAAGAACTGTCAGCCAGAGCAAGCTGTGGATCCGAGCAG- 3349
QY 3363 ATGCTTTTCAAGGCCCAAGAGCAGCAGTAAATGATGATCAGAAAGCAGGATTAATATGGAC 3422
Db 3349 -----TTCGAGTTCAGGAGTCCAGCATGATGCTGAGCAGAAAGCCCTGTTTTCCCAC 3403
QY 3423 AGACATACCCAGCAGCAGCGGCTCCAAATGCAAGAGGCTTTTCATCTTCAAGGCAATAC 3482
Db 3404 AGCAGTACGCTCTCAGACACAATATGGCCAGGGAGCTATAATCCATGCAAGATCCA- 3463
QY 3483 CATCTTTTAACTCTATGATGAATCAGATGAACCAAGCAAGCAATTTCTCTCTCCAGGAA 3542
Db 3463 -----AACTTCCACACCTGGACAGCGGCAAAATTTATCCACACTCCGAA 3508
QY 3543 TGCACCCAGCAGCAACATCATGAGACCCCGGACA-----AACACCCCAAGCAACTTA 3596
Db 3509 TGCAGCCCGCGCGGCGCTCAGGCCACACAGGCAATGTCAGAACCCAGCCCAACACTGA 3568
QY 3597 GAATGAGCTTTCAGCAGAGCTGTCAGGGCCAGCAGTTTTTGAATCAGAGCCGACAGGCAC 3656
Db 3569 GACTTCAGCTCCAACTCGCTCCAAAGCAG-----CAGAACCCGCGCCGC 3616
QY 3657 TTGAATTTGAAAATGAAAAACCCCTACTGCTGGTGTGCTGCGGTGATGAGGCTTATGATGC 3716
Db 3617 T---AATGAACAGATCAGCGCGTTTCCAAATGTGAATTTGACTTTGAGGCTTGGCGTGC 3673
QY 3717 AGCCCGCAGCAGGTTTCTTAATGCTCAATGCTGCCCAACGACAGCAGAGCTGCTAA 3776
Db 3674 CAACT---CAGGCTCTATCAATGACAGATGCTGCCCGCAGAGGAGGAAATCTTA 3730
QY 3777 GTCATCACTTCCGACAAACAGAGGTTGGCTATGATGATGACAGCAGCAACACAGCAGC 3836
Db 3731 -----ACCAGCATCTCCGCGAGA 3748
QY 3837 AGCAGCAGCAGCAGCAGCAACAGCAACAGCAGCAACAGCAGCAACCCAGGCT 3896
Db 3749 GACAGATGATCAGCAGCAGCAGGTCAGCAGCGAATTTGATGATGAGAGCAGAGGT 3808
QY 3897 TCAGCCCACTCTCAATGACTGCTTCCCGCAGCATGATGGCTTTTGGCAGGACCCA 3956
Db 3809 TGAATATGACCCCAAGCATGTTGCTCCCACTGCTTTTACCAGCAGCATGACGAAACCC 3868
QY 3957 CAATGCCAAGCTCTCCGCAACAGTTTCCATATCAACCAAAATTTGATGGGCAAC 4016
Db 3869 GGATCCCGCAAGCCAAATGCCAGCAGTTTCCATTTCTCCGAACTACGGAATAAGTCAGC 3928

```
QY 4017 AACCATCCAGCCCTTGGTCGAGTGTCTAGTCTCTCCCAATGCAATGATGTCGTCAGAA 4076
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3929 AACCTGATCTCGCTTACTGGAGCTACAACTCCCGAGAGTCTCTAATGTCCTCCCGA 3988
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4077 TGGTCCCTCCAGAAATCCCATGATGCAACACCCGAGC---TGCATCCATCTATCATG 4133
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3989 TGGCACAATACGAGAGTCCCATGATGTCAGCAGCTCTCAGGCCAACCCAGCACCAGCCG 4048
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4134 CCTCAGAATGAAGGCTGGCCATCAGGAATTTTGGCCAGGAACAGCTCTTTTCCAGC 4193
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4049 CTTACAGATTAACGATGGGCAGAGGAGCATGGCGGAAACAGCATGTTCTCACAAC 4108
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 4194 AGCAGTTTCCCGACGAGGGAATCTCGAGTGTATAGTATG 4234
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4109 AGTCCCCACACACTTTGGGCAACAGCGAACAACAGCATG 4149
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AF000582
US Mus musculus nuclear receptor coactivator protein 2 mRNA 25-JUN-1997
ORIGIN
cda
ACCESSION AF000582
NID 92213816
VERSION AF000582.1 GI:2213816
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Torchia, J., Rose, D.W., Inostroza, J., Kamei, Y., Westin, S.,
Glass, C.K., and Rosenfield, M.G.
The transcriptional co-activator p/CIP binds CBP and mediates
nuclear-receptor function
Nature 387 (6634), 677-684 (1997)
JOURNAL
MEDLINE 97336097
REFERENCE 2 (bases 1 to 4771)
AUTHORS Torchia, J. and Rosenfield, M.G.
DIRECT SUBMISSION
Submitted (21-APR-1997) Medicine, Howard Hughes Medical Institute
at the University of California, San Diego, 9500 Gilman Drive,
C.M.M. 345, La Jolla, CA 92093-0648, USA
LOCATION/Qualifiers
1. 4771
/organism="Mus musculus"
/db_xref="taxon:10090"
152..4543
/product="nuclear receptor coactivator"
/codon_start=1
/protein_id="AAB61575.1"
/db_xref="PIR:G2213817"
/db_xref="GI:2213817"
/translation="MSGMGENTDPSRAETRRKCEPDQLGPSRKSTERNRQENK
YIEELAEIIFADINDFNFKPKDCKALLKTVKIQIEOEKAAANDEVOKSDV
SSTGGVIDKDALGPMLEALDGEFFVNVLEGSVVFERNVTVLYRNQELMKSIV
SILVAGDTEFVNKLLPKSMVNGSGWGEPPRRSHFNCMLVKPLPLDSEEGHDSQ
EAKQIEAWQCFVSPKSIKEEGEDLQSLIVVHDEPHGKTNSSLRKLYHPGPP
RODFTGHYHREAGLRSGKDAFRSSTHSMKGLYHMPRRHHVLRQGLAFS
QIYRFLSDGLVLAQTKSLIRQTTNEPOLVLSLHMLHREQNVCMNPDLTGQAMG
KPLNPISSSSPHALDCSGNPQDWTLSNINFPNMPKQEQMGPMKTRFVGSGGMNV
SGMAOTPOGSNYALKWNSPQSSPGMNPQOASVLSLPRMSPGVAGSPRIPEPSQ
PAGNLHPVGVCSSTGSHSYNSLSNLAQLSSEGHVSLGSSLASFDLMKGNLQNSP
VNMNPPPLSGKSLDSDCFGLYGEPSKGTQAEASCHPKPKKQPNDSMPQASGD
RAEGHRLHLSKGTQLQLLTTSKQMEPSPLPSLSLTDNKDSTGLPGSPHGTFS
LKEKHLHRLLODSSPVDLAKTAETAKELSESSTAPGSEVTVKQEPAPFKK
ENALLRLDKDDTLDLQSLPDKLPDTPAGTSPKQALINDLMOLTADSSPSPA
PGSELDNLEETLDLQSLPDKLPDTPAGTSPKQALINDLMOLTADSSPSPA
GAOKAALCMSOSSFNNPPGOLGRLPYQNLPLDITLQSPGAGFPFPIRNSPYVI
PDQSMGNMGOGSGNGLNNSGTSRSPMSGEWAPQSTCESTLVLPLVP
RTDUSKEARFNPTASIPMGANSOLQROMLQSVNMNIGPEEMNMGPGYNOQAP
PNOTAPWPESTLIPDQASFASONRQPFSSPDLLCLPHPAESPSDEGALLDLYAL
```

```

RNFDELLEIDRALGIPELVSQAVDAEOLFSSOESSIMLEKQKPPVPOQYASQAQMAQ
GGVNPQDPNFHTMGORPNYTTLRMPRGLRPTGIQONOPNLRQLORLQAOQNR
OPLMNOISSVSNVNLRLPGVPTQAPINAOMLAQORREILNOHLRQOMQOYQOETL
MQGQSLNVTIPSVVAPAGLPAAMSNPRIPOANAQQPFPNPNYGISQDPDGFGTATP
QSLMSPRVAHTQSPMWOQSQANPAYQPTSDMNGWAQSGMNGSMFSSQSPPHFGQQA
NTSMYNNMNIYSVMATNTGLSSMNQMTGQNSMTSVTSVPTSLGSPMGPEQVNDPAL
RGNLFPNQLLGMIMIKQEGDASRKYC"
BASE COUNT 1292 a 1347 c 1228 g 904 t
ORIGIN
Query Match 10.5%; Score 474; DB 12; Length 4771;
Best Local Similarity 50.5%; Pred. No. 3.8e-103;
Matches 2180; Conservative 0; Mismatches 1845; Indels 289; Gaps 31;
QY 64 CAGTTGCTGATGATATTAAGATGAGTGGATTAAGGAGAAACCTTGGATCCACGTGGCCAG 123
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 CAGTTGCTGATGATGTTCAAGATGAGTGGATGGGAGAAACACCTCTGACCGCTCCAG 189
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 124 T---GATTACGAAACGCAAAATTCCTATGTACTCCAGGCAAGGTCTTACTCTGCAG 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 GCGAGAGACCAAAACGCAAGGAATGTCCGACACAGCTCGGACCCAGCCCCAAAGAG 249
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 TGTGAAACGAGAGCGGAGCAGGAGAAATATATATTGAAGAATTGGCTGAGCTGAT 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 CACTGAGAACGGAACCGCAGCAGGAGGAGAAATAGTACATACAGGAGCTGGCCGAGCTGAT 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 ATCTGCCAATCTTAGTATATTTGACAAATTTCAATGTCAACCAGATAAATGTGGGATTT 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 CTTGCAAACTTTAATGATATTTGACAACTTCAACTTCAAACTTCAAAATGTGCATCCT 369
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 AAAGAAACAGTAAAGACAGATACGTCAATATAAAGAGCAAGGAAA---AATATTTCAA 357
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 AAAAGAACTGTGAAGCAGATCCGCCAGATCAAGAGCAAGAGAAAGCAGCAGCTGCCAA 429
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 358 TGATGATGATGTTTCAAAAGCCGATGATCTTCTACAGGGCAGGAGGTATTGATAAGA 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 CATAGTGAAGTGCAGAAGTCAGATGTGCTGCCACGGGCGAGGTGTCTATCGACAAGA 489
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 CTCCTTAGGACCGCTTTTACTTTCAGGCATTTGGATGGTTTCTTATTTGGTGAATCGA 477
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 TGCATGGGGCCATGATGCTTGAAGCCCTCGATGGGTCTTCTTCTGTTGTAACCTGGA 549
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 478 GGCAACATCTGTTTGTATCAGAAAATGTCAACAATACCTGCAATATAAGCAAGAGA 537
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 AGGAGCTGTGTTGCTGTGTCAGAAATGTGACACAGATATCTACGGTATAACCAAGA 609
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 538 CTGGTTAACACAAGTGTTTACAATATCTTACATGAAGAGACAGAAAGGATTTTCTTAA 597
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 GCTGATGAACAAGAGTGTCTACAGCATCTCGATGTGCGGGACCACTGAATTTGTCAA 669
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 GAA---TTTACCAAAATCTACAGTTAATGAGTTTCTTGGACAATAGCCCCAAAGACA 654
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 GAACCTGTGTCAAAAGTCCATGGTGAATGGAGATCTGCTGTGGAGAACCTCCAGCG 729
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 655 AAAAGCCATACATTTAATTCCTGATGTTGATCAAAAACACCATATATCTTGGAGA 714
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 730 GAGCAGCATACCTTCAACTGTCTGATGCTGGTGAAGCCTTTGCGAGCTTCAGAAAGAG 789
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 715 CATAAACCCAGTCTCTGAAATGCCCCAGAGATATGAACAATGACAGTCTTTGGCCCTG 774
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 790 AGGCCATGATAGCCAGGAAGCCCATCAGAAATACGAGCGGATGAGTCTGCTGCTGTC 849
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 775 TCAGCCAGCAGCTATGATGAGGAGGAGGAGAGATTTGCAATCTTGTATGATCTGTGGC 834
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 850 TCAGCCCAAGTCCATCAAGAGGAGGAGGAGAGATTTGCAAGTCTGCTGCTGAT-TGTGTC 908
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 835 ACGGCGCATTTACTCAGGAGAGAAAGAAATTTCCATCAAACTCAGAGCTTTTATTTACCA 894
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 909 ACGAAG-ATCCCCATGAAGGAGAAACCACTCTTCCCTCATCAGAAAGCTTTTACCACCG 967
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 895 ACATGATCTTTTCAGGAAAGGTTGCAATATAGATACAAATTTACTGAGATCCTCCATCAG 954
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


MAGGNTNVQNNPNPMAQMQMSSLQMPGMNTVCPEQINDPALRHTGLYCNQLSSTDLLK
TEADGTCDDKKTEFFFSVTTD"

mat_peptide	205	.4395			
BASE COUNT	1462 a	1165 c	1012 g	1082 t	
ORIGIN					

Query Match 7.3%; Score 328.4; DB 10; Length 4721;
Best Local Similarity 59.7%; Pred. NO. 2.4e-68;
Matches 633; Conservative 0; Mismatches 406; Indels 21;

Qy	156	CTCCAGACAAAGTGCTTTACCTGAGTGGTGAAAAAAGGAGACGGGACAGGAGAAAGTAAAT	215
Db	266	CGCCATGTGACACACTGGCATCAAGCACGGAAGAGAGCGCAGGAGCAGAGAAATAAAT	325
Qy	216	ATATTGAAGAAATTTGGCTGACGTGATATCTGCCAATCTTAGTGATATTGACAATTTCCAATG	275
Db	326	ATTTAGAAGAACTAGCTGAGTACTGTCTGCCAACATTAGTGACATTTGACAGCTTGAGTG	385
Qy	276	TCAAACGAGATAAATGTGCGATTTTAAAGGAACACAGTAAGACAGATACGTCNAATAAA	334
Db	386	TAAACCAGACAAATGCAAGATTTTGAAGAAAAACAGTCGATCAGATACAGCTAATGAAGA	445
Qy	334	-AGAGCAAGGAAANAACATATTTCCAATGATGATGTTTCAAAGCGCGATGTATCTCTA	392
Db	446	GAATGGAAACAAGAGAAATCAACAACATGTAGCAGCTATACAGAAATCAGACATCTCATCA	505
Qy	393	CAGGGCAGGGAGTATTGATAAAGACTCCCTAGGACCGCTTTTACTTCAGGCATTCGGATG	452
Db	506	GTAGTCAAGGAGTATAGAAAAGGAATCCTTTGGGACCTCTTCTTTGGAGGCTTTGGATG	565
Qy	453	GTTTCTTATTTGGTGGAAATCGAGAGGCAAAACATGTATTTGTATCAGAAAAATGTCACAC	512
Db	566	GATTTTCTTTTGTGTGAACGTGTCAAGGGAGAAATGTAATTTGTGTGTCAGAGAATGTAACCA	625
Qy	513	AATACCTGCAATATACGACAGAGGACCTGGTTAACACAGTGTTTACAATATCTTACATG	572
Db	626	GCTACTTAGTTTACATCAGGAGGAATTAATGAATACCAGCGCTTCACGCATACTGCACG	685
Qy	573	AAGAAGACAGAAAGGATTTTCTTAAAGAAAT---TTACCAAAATCTACAGTAAATGAGATTT	629
Db	686	TGGGGATCATGCAGAAATTTGTGAAGAACTCTGCTACCAAAATCACATAGTAATGGATTC	745
Qy	630	CTTGGACAAATGAGCCCCAAAGACAAAAAGCCATACATTTAATTCGCGTATGTTGATGA	689
Db	746	CTTGGCCCTCAAGAGGCAACACGACGAAATAGCCATACCTTTAACTCGAGGATGCTAATTC	805
Qy	690	AAACACCAATGATATTTCTGGAGACATAAACGCCAGTCTGAAATGCGCCAGAGATATG	749
Db	806	ACCTCCAGATGACCCAGG------ACCAGAACCAAGAAAGCTTGCACGCTTTG	856
Qy	750	AAACAATGCAGTGCTTTGGCCCTCTCTCAGCCAGCGCTATGATGGAGGAAGGGGAAGATT	809
Db	857	AAGTAATGCAGTGTTTCACTGTGTACAGGCCAAATCAATTCAGAGGATGGAGAGATT	916
Qy	810	TGCAATCTTGATGATCTGTGTGGCAGCCGCGATTACTACAGAGAGAAAGACATTTCCAT	869
Db	917	TCCAGTCATGCTGATTTGATTGACGGCGGATTTACTCTGGCCCTC-----CAGCTATTA	970
Qy	870	CAAAACCTGAGAGCTTTATPACCAGACATGATCTTTCAGGAAAGGTTGTCAATATAGATA	929
Db	971	CGGGTGTAAGTCTTTATGACCAACGAAGATACTACAGGTAATATCTCTATTGATA	1031
Qy	930	CAAAATTCAGTATCCTCCATGAGGCCCTGGCTTTGAAGATATAATCCGAAGGTGATTTC	989
Db	1031	CTAGTTCCTTCAGAGCTGCTGGCAGAACTGGTTGGGAAGATTTAGTGAGGAAGTGCATTT	1091
Qy	990	AGAGATTTTTTAGTCTAAATGATGGGCACTGATGGTCCCGAAGACCTCACTATCAGAGAG	1041
Db	1091	ATGCTTTTTTCCAACTTCAGGGCAGAGAACCATCTTATGCGACACAGCTGTTCCTCAAGAG	1151
Qy	1050	CTTATCTTAATGCCCATGCAGAAACCCCAAGTATATGGATTCTCGTTGGGTGATGGAACTA	1101

Db	1151	TGAGTACTGCTGGC	ACTGCTCCAGCCCTCTCTATAGTATCATATNTGAATGATGGGACAA	1210
Qy	1110	TAGTGACTGCACAGACAAA	AAAGCAAACTCTTCCGAAATCTGTAAACAAATGATCGACATG	1169
Db	1211	TGCTTAGCGCCCAACACAA	AGTGTAAACTTTGTCTACCTTCAAAGTCCAGACATGCAACCTT	1270
Qy	1170	GCTTTGTCTCAACCCCACTT	CTCTTCAGAGAGACAAATGG	1209
Db	1371	TCATCATGGGAATTCATAT	CATCTGACAGGGAGCAGATGG	1310

RESULT 15

HSU90661	HSU90661	4547 bp	mRNA	PRI	25-MAR-1997
LOCUS	Human steroid receptor coactivator-1				
DEFINITION	Human steroid receptor coactivator-1 mRNA, complete cds.				
ACCESSION	U90661				
NID	91906027				
VERSION	U90661.1	GI:1906027			

KEYWORDS: human
SOURCE:

SOURCE	ORGANISM
Human.	Homo sapiens
Mammal.	Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

AUTHORS Spencer, T. E., Jenster, G., Onate, S., Tsai, M.-J. and O'Malley, B. W.
TITLE Cloning of full-length human steroid receptor coactivator-one (SRC-1)

JOURNAL Unpublished (1997)

REFERENCE	AUTHORS	TITLE
2 (bases 1 to 4547)	Spencer, T. E., Jenster, G., Onate, S., Tsai, M.-J. and O'Malley, B. W.	Direct Submission

JOURNAL Submitted (24-FEB-1997) Cell Biology, Baylor

FEATURES

CDS

```

/note="SRC-1"
/codon_start=1
/product="steroid receptor coactivator-1"
/protein_id="AAB50242.1"
/db_xref="PID:g1905028"

```

/translation="MSGLGDSDDPANPSHKKRKGPCDPTLASTEKRRREQENKYLE
 ELAELLSANTSDLSLWPKDKILKTKVDQOLMRQOEKSTDDDDVOKSDISS
 SOQVIEKESGLPILLEALGDOFFVFNCEGRISYFVENWTSYLGYNQELMNTSVYAIL
 HVGDHAEFVKNLPLKPSLVNGVPMWQATRNHSHTNCRMLIHPDPGPTQNBACORY
 ENVOCTVSOPKSIOBEDGEDFQCLICIAARRLPAPAITQVFSMTKQDITGRIISD
 TSSLAAGRTGWEDLVKRCIYAFQOGRPSYAROLFQEVMTGTATSPSFIILND
 GTWLASHTKCLCYPOSMDOPFIMGIHIIIDREHSGLSPODDTNSGMSI PRVPSVNP
 SISPAHGVARSSLTSPSNSNMWSTRINRQOSSDHLHSSHSNSNSOGSGFCGPGSOIV
 ANVALKQSSLOPPSSUNLNNPMEGTGISLAQFMSPPRQVTSGLATPRMPNNSF
 PPNITSISPVGNTSACNNNRASYINPIVTSIQGNNEGPNNSVGSFASPSYORQMS
 QNSPRLNTQOAKESCKNKEIASTLNEMIQSDNSDGKPLDSGLLHNDRUSDGS
 KYSOTSHKLVOLLTHTAEQOLRHADITDCKDVLCTGTTNSASANSOGSCSPSSHSS
 LTRARKILHRLQEGSPDITLUSVEPDKDSASTSVYTVQVQGNSSKLEUDASKK
 KESKDHLIRLIDKDEQDQLLTLEKAAQLPGCEITRDMGCAVTSTIKELIPASLOS
 IKLEAQOSTADLODQLLTLEKAAQLPGCEITRDMGCAVTSTIKELIPASLOS
 ATARPTSRLNRLPELELAIDNGOGPGTGDOIPTWNTNTVTAI.NQSKSEDOCISSOLD
 ELLCPTTVTEGRNDEKALLEQLVSLSGKDETELAELDRALGIDKLKVGQGLDVLSER
 FFPQQTPLIMEERPNLSQYSFPPFTANLPSPQGMVWRKPSLGTMPVTPPRG
 AFGQGMGMPROTLPNPPAQNGLRLQOOLQOQOQLIQHNRAQILNQVFAATPYGI
 NMRSGMQOQTPPLPNAQOQOQRELYSQOHRQQLIQOORAMLQROSGFNNLP
 SSGLVGTGNPLPOCAQOQOFPYPPNYGTPGTPPASTSPFSOLANPRASLANRSM
 VSRGNTGTGGQGTGQINQMQUQVQFGAGMVGQGEANFARSLSPGSSMWPMPIPP
 PQSSLNQTTPASGYSQDMKAWQOAGLNNNVTSQAQVQNQTPPGVYNNMSITLS
 MAGGNTNQNMNPMQAQMSLQPMGMTVCPQEQINDPALRHQGLKYNQSLDILLK
 TEADGTQVQOQVQVADVQCTVNLVGGDPTLNQPGPLGTQKKTSGTPQTPQAQCKSLR
 OLI,TF"

BASE COUNT	ORIGTN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

M P S R E H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
Search_pp protein - protein database search, using Smith-Waterman algorithm
on: Fri Sep 17 21:06:51 1999; MasPar time 56.90 Seconds
996.458 Million cell updates/sec
Tabular output not generated.

Title: >US-09-041-994-2
Description: (1-1415) from US09041994.pep
Perfect Score: 9849
Sequence: 1 MSGLGENDPLASDSRRKL.....MNMNPMMPMGMPGDKYK 1415
Scoring table: PAM 150
Gap 11
Searched: 122810 seqs, 40068593 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: pir60
1:pir1 2:pir2 3:pir3 4:pir4
Statistics: Mean 55.130; Variance 137.576; scale 0.401

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	pred. No.
1	9815	99.7	1417	2	T03443	receptor-associated c
2	9778	99.3	1420	2	T03749	probable nuclear rece
3	643	6.5	1061	2	A57620	steroid receptor coac
4	232	2.4	848	2	S59514	aryl hydrocarbon rece
5	239	2.4	853	2	S58375	aryl hydrocarbon rece
6	222	2.3	313	2	S07924	alpha/beta-gliadin pr
7	225	2.3	339	1	1WH028	transcription initiat
8	225	2.3	344	2	L60128	transcription factor
9	230	2.3	357	2	S18235	omega scapalin precurs
10	229	2.3	357	2	S18236	omega scapalin precurs
11	230	2.3	644	2	S39356	transcription factor
12	227	2.3	758	2	S54322	hypothetical protein
13	231	2.3	805	2	A46266	aryl hydrocarbon rece
14	226	2.3	905	1	RGBVS5	regulatory protein SN
15	231	2.3	4957	2	T03455	ALR protein - human
16	231	2.3	5262	2	T03454	ALR protein - human
17	221	2.2	609	2	A49839	odd-paired - fruit fl
18	217	2.2	2175	2	S03170	homeotic protein cut
19	208	2.1	139	2	A26992	Mopa box protein - mo
20	210	2.1	360	2	S50830	Machado-Joseph diseas
21	205	2.1	467	1	A49377	involucrin - mouse
22	202	2.1	539	2	S57972	hypothetical protein
23	204	2.1	583	2	JC5404	brain and muscle Ah r
24	202	2.1	583	2	JC5404	brain and muscle Ah r

24	204	2.1	626	2	JC5405	brain and muscle Ah r	1.66e-10
25	205	2.1	648	1	JQ1150	protein kinase (EC 2.	1.23e-10
26	204	2.1	708	2	A53185	G-box-binding factor	1.66e-10
27	203	2.1	966	2	S25365	CYC8 protein - yeast	2.24e-10
28	204	2.1	1012	2	I53172	RAE-28 - mouse	1.66e-10
29	204	2.1	1081	2	S66736	transcription activat	1.66e-10
30	202	2.1	2038	2	A43742	female sterile homeot	3.02e-10
31	210	2.1	2414	2	A54277	transcription adaptor	2.74e-11
32	205	2.1	2440	2	S39162	CREB-binding protein	1.23e-10
33	204	2.1	2441	2	S39161	CREB-binding protein	1.23e-10
34	205	2.1	3144	2	A46068	Huntington disease-as	1.66e-10
35	199	2.0	181	2	JC5406	brain and muscle Ah r	7.37e-10
36	199	2.0	238	2	JC5407	brain and muscle Ah r	7.37e-10
37	196	2.0	271	2	PC4288	brain and muscle Ah r	1.79e-09
38	198	2.0	286	2	S07923	alpha/beta-gliadin pr	9.91e-10
39	200	2.0	326	2	D22364	alpha/beta-gliadin pr	5.47e-10
40	198	2.0	398	2	S21883	DNA-binding protein p	9.91e-10
41	200	2.0	613	2	S27770	hypothetical protein	5.47e-10
42	199	2.0	626	2	JE0270	Arnt-like PAS protein	7.37e-10
43	197	2.0	737	2	S47857	basic protein, cyto	1.33e-09
44	200	2.0	756	2	S60966	probable protein kina	5.47e-10
45	196	2.0	813	2	JC5809	HIF-1 alpha - rat	1.79e-09

ALIGNMENTS

RESULT 1
ENTRY T03443 #type complete
TITLE receptor-associated coactivator 3 - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Mar-1999
ACCESSIONS T03443
REFERENCE Z14950
#authors Li, H.; Gomes, P.J.; Chen, J.D.
#journal Proc. Natl. Acad. Sci. U.S.A. (1997) 94:8479-8484
#title RAC3, a steroid/nuclear receptor-associated coactivator that is related to SRC-1 and TIF2.
#accession T03443
#status preliminary; translated from GR/EMBL/DBJ
#molecule_type mRNA
#residues 1-1417 ##label LIH
##cross-references EMBL:AF010227; NID:g2318005; PID:g2318006
##experimental_source cell line HeLa
GENETICS
#gene RAC3
#summary RAC3
#length 1417 #molecular-weight 154535 #checksum 7159

SUMMARY									
Query Match	99.7%	Score	9815	DB 2:	Length	1417			
Best Local Similarity	99.7%	Pred. No.	0.00e+00						
Matches	1413	Conservative	1	Mismatches	1	Indels	2	Gaps	1
Db	1	MSGLGENLPLASDSRRKLPCDTPQGGLTSCGEKRRRQESKYIEELAEISANLSDID	60						
QY	1	MSGLGENLPLASDSRRKLPCDTPQGGLTSCGEKRRRQESKYIEELAEISANLSDID	60						
Db	61	NFWVKPKCAILKETVQRIQIKQEKTTISNDDVDQKADVSSTGGQVVIDKDSGLPLLQA	120						
QY	61	NFWVKPKCAILKETVQRIQIKQEKTTISNDDVDQKADVSSTGGQVVIDKDSGLPLLQA	120						
Db	121	LDGFLFWNREANIVFVSENVTOYQYKQEDLVNTSYNIIHEEDRKDFLKNLPKSTVNG	180						
QY	121	LDGFLFWNREANIVFVSENVTOYQYKQEDLVNTSYNIIHEEDRKDFLKNLPKSTVNG	180						
Db	181	VSWTNETQKSHFTFCRMLKTPHDILEDINASPENRQRYETMQCFALSQPRAMEE	240						
QY	181	VSWTNETQKSHFTFCRMLKTPHDILEDINASPENRQRYETMQCFALSQPRAMEE	240						
Db	241	DLQSCMICVARRITTCERTFPSPESFIRHDLGKGVNIDTNSLRSSMRPGFEDIIRRC	300						
QY	241	DLQSCMICVARRITTCERTFPSPESFIRHDLGKGVNIDTNSLRSSMRPGFEDIIRRC	300						
Db	301	IQRFFSLNDQSQSKRRHYQAYLNGHAETPVYRFSIADGTIVTAQTKSLFRNPVTNR	360						

```
QY 301 IORFFSLNDGQSWQKRHYQAYLNGHAETPVYRFLADGTTVTAQTKSLFRNPVTNDR 360
Db 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSNQGLOMPSSRAYGLA 420
QY 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSNQGLOMPSSRAYGLA 420
Db 421 DPSTTGQSGARYGSSNIASTLTPGQSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
QY 421 DPSTTGQSGARYGSSNIASTLTPGQSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
Db 481 MISPRNRGSPKIASHQFSPVAGVHSPMAGSNTGNHSSSSLSALQAISEGVGTSLST 540
QY 481 MISPRNRGSPKIASHQFSPVAGVHSPMAGSNTGNHSSSSLSALQAISEGVGTSLST 540
Db 541 LSSPGPKLDNSPNMNIQPSKVSQDSKSPGLFYCDQNPVSSCMQSNRDLHLSKESKE 600
QY 541 LSSPGPKLDNSPNMNIQPSKVSQDSKSPGLFYCDQNPVSSCMQSNRDLHLSKESKE 600
Db 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSLSNPLDSSCKESSVSVTSPSGVS 660
QY 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSLSNPLDSSCKESSVSVTSPSGVS 660
Db 661 STSGVSVSTSNHGSLLQEKHRIHLKLLQNGNSPAEVAKITAETATGKDTSSITSCGDGN 720
QY 661 STSGVSVSTSNHGSLLQEKHRIHLKLLQNGNSPAEVAKITAETATGKDTSSITSCGDGN 720
Db 721 VYKQQLSPKKENALLRYLLDRDDPSDALSKELOQVQEGVDNKMSTSTSISSSOE 780
QY 721 VYKQQLSPKKENALLRYLLDRDDPSDALSKELOQVQEGVDNKMSTSTSISSSOE 780
Db 781 KDPKIKTETSEEGSDLDNLDAILDGTSDFYNNSSISNGSHLTKQVFGTNSLGLK 840
QY 781 KDPKIKTETSEEGSDLDNLDAILDGTSDFYNNSSISNGSHLTKQVFGTNSLGLK 840
Db 841 SSQSVQSRPPYNNRAVLDSPVSGSPVPKNIAPFPLPKPMLGGNPRMMDSENGYS 900
QY 841 SSQSVQSRPPYNNRAVLDSPVSGSPVPKNIAPFPLPKPMLGGNPRMMDSENGYS 900
Db 901 SMGGNRRNVTVTQTPSSGDWGLPNKAGMEPWNNSMGRPGDYNTSLPRALGGS IPT 960
QY 901 SMGGNRRNVTVTQTPSSGDWGLPNKAGMEPWNNSMGRPGDYNTSLPRALGGS IPT 960
Db 961 LPLRSNSIPGARVPVLQOQOQMLQMRPEITPMGNGANPYCQAAASNQLGSWPDGMLSMQV 1020
QY 961 LPLRSNSIPGARVPVLQOQOQMLQMRPEITPMGNGANPYCQAAASNQLGSWPDGMLSMQV 1020
Db 1021 SHGTQNRPLRLNSLDLVGPPSNLQSGSDERALLDQLHTLLSNTDQATGLEEIDRALGIPE 1080
QY 1021 SHGTQNRPLRLNSLDLVGPPSNLQSGSDERALLDQLHTLLSNTDQATGLEEIDRALGIPE 1080
Db 1081 LYNQGALEPKQDAFQGGQEAAYMDQKAGLYGQTYPAQGPMPQGGFHLQGGSPSFSNMN 1140
QY 1081 LYNQGALEPKQDAFQGGQEAAYMDQKAGLYGQTYPAQGPMPQGGFHLQGGSPSFSNMN 1140
Db 1141 QMNOQGNFPLQGMHPRANIMRPTNPKOLRMQLQRLQGGQFLNQSROALBLKKNENPTA 1200
QY 1141 QMNOQGNFPLQGMHPRANIMRPTNPKOLRMQLQRLQGGQFLNQSROALBLKKNENPTA 1200
Db 1201 GGAAYVRPMQVQOQFLNAQYVAQRSRELLSHHFQQRVAMMMQOQOQOQOQOQOQOQ 1260
QY 1201 GGAAYVRPMQVQOQFLNAQYVAQRSRELLSHHFQQRVAMMMQOQOQOQOQOQOQOQ 1260
Db 1261 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1320
QY 1261 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1320
Db 1321 VSSPPNAMSSRMGFSQNPMMHQPAASTIQSSEMKGWP SGNLARNSSFSQOQFAHQGNP 1380
QY 1319 VSSPPNAMSSRMGFSQNPMMHQPAASTIQSSEMKGWP SGNLARNSSFSQOQFAHQGNP 1378
Db 1381 AVYSVHMNGSSGHGMQNMNPMGMPGPDQKYC 1417
```

```
QY 1379 AVYSVHMNGSSGHGMQNMNPMGMPGPDQKYC 1415

RESULT 2
ENTRY T03749 #type complete
TITLE probable nuclear receptor coactivator - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
ACCESSIONS T03749
REFERENCE Z15053
#authors Tanner, M.M.; Guan, X.Y.; Sauter, G.; Kallioniemi, O.P.;
#journal Science (1997) 277:965-968
#title AIB1, a steroid receptor coactivator amplified in breast and
#accession T03749
##status preliminary; translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-1420 #label ANZ
##cross-references EMBL:AF012108; NID:g2331249; PID:g2331250
GENETICS
#gene AIB1
#map_position 20
SUMMARY #length 1420 #molecular_weight 154891 #checksum 9406

Query Match 99.3%; Score 9778; DB 2; Length 1420;
Best Local Similarity 99.4%; Pred. No. 0.00e+00;
Matches 1411; Conservative 3; Mismatches 1; Indels 5; Gaps 1;

Db 1 MSGLCENLDPLASDRKRLPCDTPGQGLTCSGKRRREQESKYEELAEILSANLSDID 60
QY 1 MSGLCENLDPLASDRKRLPCDTPGQGLTCSGKRRREQESKYEELAEILSANLSDID 60
Db 61 NFNVKPKCAILKETVROIQIRKEQKTIISNDQVQADSVSTGGQVTDKDSLGLPQLQA 120
QY 61 NFNVKPKCAILKETVROIQIRKEQKTIISNDQVQADSVSTGGQVTDKDSLGLPQLQA 120
Db 121 LDGFLFVYVNRGNIVFVSENVTQYLOKQEDLVNTSVYNIILHEEDRKDFLKNLPKSTVNG 180
QY 121 LDGFLFVYVNRGNIVFVSENVTQYLOKQEDLVNTSVYNIILHEEDRKDFLKNLPKSTVNG 180
Db 181 VSWNETORQKSHFNCRMLAKTTPHDILEDINASPENRQRYETWCQFALSPRAMMEGE 240
QY 181 VSWNETORQKSHFNCRMLAKTTPHDILEDINASPENRQRYETWCQFALSPRAMMEGE 240
Db 241 DLQSCMICVARRITTGERTFFSPNPESFTRHDLGSKVYVNIIDTNSLRSSMRPFEDIIRRC 300
QY 241 DLQSCMICVARRITTGERTFFSPNPESFTRHDLGSKVYVNIIDTNSLRSSMRPFEDIIRRC 300
Db 301 IORFFSLNDGQSWQKRHYQAYLNGHAETPVYRFLADGTTVTAQTKSLFRNPVTNDR 360
QY 301 IORFFSLNDGQSWQKRHYQAYLNGHAETPVYRFLADGTTVTAQTKSLFRNPVTNDR 360
Db 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSNQGLOMPSSRAYGLA 420
QY 361 HGFVSTHFLQREQNGYRPNPNVGGIIRPPMAGCNSSVGGMSNQGLOMPSSRAYGLA 420
Db 421 DPSTTGQSGARYGSSNIASTLTPGQSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
QY 421 DPSTTGQSGARYGSSNIASTLTPGQSPSSYQNNYGLNMSPPHPSGLAPNQNI 480
Db 481 MISPRNRGSPKIASHQFSPVAGVHSPMAGSNTGNHSSSSLSALQAISEGVGTSLST 540
QY 481 MISPRNRGSPKIASHQFSPVAGVHSPMAGSNTGNHSSSSLSALQAISEGVGTSLST 540
Db 541 LSSPGPKLDNSPNMNIQPSKVSQDSKSPGLFYCDQNPVSSCMQSNRDLHLSKESKE 600
QY 541 LSSPGPKLDNSPNMNIQPSKVSQDSKSPGLFYCDQNPVSSCMQSNRDLHLSKESKE 600
Db 601 SSVGAENQRPGLSKGHKKLLQLLTCSSDDRGHSSLSNPLDSSCKESSVSVTSPSGVS 660
```

```
QY 601 SSVGAENQRPLESKGKHLQLLLTSSDDRHSSLTNSPLDSSCKESSVSVTSPGVS 660
Db 661 SSTGGVSSSTNMGSLLOEKHRLHLKLLQNGNSPAEVAKITABATKDKTSITSCDGN 720
QY 661 SSTGGVSSSTNMGSLLOEKHRLHLKLLQNGNSPAEVAKITABATKDKTSITSCDGN 720
Db 721 VVKEQLSPKKENALLRYLLDRDDPSDALSKELQPOVEGVNKMSSCTSTIPSSOE 780
QY 721 VVKEQLSPKKENALLRYLLDRDDPSDALSKELQPOVEGVNKMSSCTSTIPSSOE 780
Db 781 KDPRIKTTSETSEGGDLNDLDAIIGDLSDFYNNSSSSNGSHLGTQQVFGQFNSLGLK 840
QY 781 KDPRIKTTSETSEGGDLNDLDAIIGDLSDFYNNSSSSNGSHLGTQQVFGQFNSLGLK 840
Db 841 SSQVQSIRPPYNRAVSLDSPVSGSPVKNISAFPMPLPKPMLGNGNPRMDSQENYS 900
QY 841 SSQVQSIRPPYNRAVSLDSPVSGSPVKNISAFPMPLPKPMLGNGNPRMDSQENYS 900
QY 901 SMGGPNRNVTVTQPSGGDGLPNKAGRMPEMNSNGRPGGDYNTSLRPALGGSIPT 960
QY 901 SMGGPNRNVTVTQPSGGDGLPNKAGRMPEMNSNGRPGGDYNTSLRPALGGSIPT 960
Db 961 LPLRSNIPGARPYLQOQOQMLQMRPEIPMGMGANPYGQAAASNLQSGWPDGMLSEQV 1020
QY 961 LPLRSNIPGARPYLQOQOQMLQMRPEIPMGMGANPYGQAAASNLQSGWPDGMLSEQV 1020
Db 1021 SHGTQNRPLRLNSLDDLVGPPSNLEGSQDERALLDQLHTLLSNTDATGLEIDRALGIPE 1080
QY 1021 SHGTQNRPLRLNSLDDLVGPPSNLEGSQDERALLDQLHTLLSNTDATGLEIDRALGIPE 1080
Db 1081 LVNGQALEPKQDAFOGAEAAVMDQKAGLYGQTPYPAQGPMPQCGFHLQSGPSFNSMN 1140
QY 1081 LVNGQALEPKQDAFOGAEAAVMDQKAGLYGQTPYPAQGPMPQCGFHLQSGPSFNSMN 1140
Db 1141 QMNOGQNFPLQGMHPRANIMPRNTPKQLRMQLQRLQGOOFLNQSRQALELKNENPTA 1200
QY 1141 QMNOGQNFPLQGMHPRANIMPRNTPKQLRMQLQRLQGOOFLNQSRQALELKNENPTA 1200
Db 1201 GGAAMRPMQPOQGFFLNAQVARSRELLSHHFRQQRVAMMMQOQOQOQOQOQOQOQ 1260
QY 1201 GGAAMRPMQPOQGFFLNAQVARSRELLSHHFRQQRVAMMMQOQOQOQOQOQOQOQ 1260
Db 1261 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1320
QY 1261 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 1315
Db 1321 FGVRSSPPNMMSSRMGSPQNMHQPAAASIOSSEKMGWPSGNLARNSSFSQOQFAHQ 1380
QY 1316 FGVRSSPPNMMSSRMGSPQNMHQPAAASIOSSEKMGWPSGNLARNSSFSQOQFAHQ 1375
Db 1381 GNPAYSMVHMNGSSGHGMQNMNPMPSGMPGPDQKYC 1420
QY 1376 GNPAYSMVHMNGSSGHGMQNMNPMPSGMPGPDQKYC 1415
RESULT 3
ENTRY #type complete
TITLE steroid receptor coactivator 1 - human
ORGANISM formal_name Homo sapiens #common_name man
DATE 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change
30-May-1997
ACCESSIONS A57620
REFERENCE A57620
#authors Onate, S.A.; Tsai, S.Y.; Tsai, M.J.; O'Malley, B.W.
#journal Science (1995) 270:1354-1357
#title Sequence and characterization of a coactivator for the steroid hormone receptor superfamily.
#cross-references MUID:96085063
#accession A57620
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-1061 #label ONA
KEYWORDS steroid hormone receptor
```

```
SUMMARY #length 1061 #molecular_weight 114166 #checksum 398
Query Match 6.5%; Score 643; DB 2; Length 1061;
Best Local Similarity 36.2%; Pred. No. 1.76e-76;
Matches 113; Conservative 89; Mismatches 89; Indels 21; Gaps 18;
Db 456 MDGAVTSVTKSEILPASLOSATARTPSRLN-RLPELELEAIDNQFGQPGTGOI-PWTN 513
QY 954 LGSISITPLRLNSIPGA-REVLIQOQOQMLQMRPEIPMGMGANPYGQAAASNLQSGWPD 1012
Db 514 NTVTAIINQSK-SEDOCI-SGQDELICPPTTVEGRNDEKALLEGVFLSGKDETELAEL 571
QY 1013 GMLSMQEVSGTGNRPLRLNSLDDLVGPPSNLEGSQDERALLDQLHTLLSNTDATGLEEI 1072
Db 572 DRALGIDKLV-QGGGLDVLSEFPFPPQOATPPLIMEERNPLNSQYSPFPPTANLPSPFQ 630
QY 1073 DRALGIPELVNGQALEPKQDAFOGQEA--VMDQKAGLYGQTPYPAQGP-P-MGGGFH- 1127
Db 631 MYRQKPSLGTMPVQVTPPRGAFSPGCMQPRQTLNRP-APNQLRLQLOQRGQOQOLI 689
QY 1128 LOGQSPSFNSMNMQNO-QGNF-PLQGMHPRANIMPRNTPKQLRMQLQRLQGOO-FL 1184
Db 690 HQNRQAI-LNOFAATAPVGINMSGMOQOITPOPPLNAQMLAORRELYSQOHRQ-R-QL 746
QY 1185 NQSRQALELKNENPTAGCAAVMRPMQPO---QGFLNAQVARSRELLSHHFRQQRVAM 1241
Db 747 IQQORAMLMRQ 758
QY 1242 MMQOQOQOQOQOQ 1253
RESULT 4
ENTRY #type complete
TITLE aryl hydrocarbon receptor - human
ALTERNATE_NAMES Ah receptor
ORGANISM formal_name Homo sapiens #common_name man
DATE 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change
10-Sep-1997
ACCESSIONS S59514; JX0356; S41124
REFERENCE S59514
#authors Dolwick, K.M.; Schmidt, J.V.; Carver, L.A.; Swanson, H.I.; Bradfield, C.A.
#journal Mol. Pharmacol. (1993) 44:911-917
#title Cloning and expression of a human Ah receptor cDNA.
#cross-references MUID:94067047
#accession S59514
#status preliminary
#molecule_type mRNA
#residues 1-848 #label DOL
#cross-references EMBL:L19872; NID:g416141; PID:g416142
REFERENCE JX0356
#authors Ema, M.; Matsushita, N.; Sogawa, K.; Ariyama, T.; Inazawa, J.; Nemoto, T.; Ota, M.; Oshimura, M.; Fujii-Kuriyama, Y.
#journal J. Biochem. (1994) 116:845-851
#title Human arylhydrocarbon receptor: Functional expression and chromosomal assignment to 7p21.
#accession JX0356
#molecule_type mRNA
#residues 1-4,'G',6-356,'H',358-460,'N',462-631,633-848 #label EMA
#experimental_source lung
REFERENCE S41124
#authors Itoh, S.; Kamataki, T.
#journal Nucleic Acids Res. (1993) 21:3578
#title Human Ah receptor cDNA: analysis for highly conserved sequences.
#cross-references MUID:93347997
#accession S41124
#status preliminary
#molecule_type mRNA
#residues 1-404,'N',406-797,'E',799-806,'FK', #label ITO
COMMENT This protein is involved in the process in which 2,3,7,8-tetrachlorodibenzo-p-dioxin exerts its effects.
```

```
GENETICS
#gene
#cross-references GDB:138471; OMIM:600253
#map_position 7p21-7p15
KEYWORDS
FEATURE
117-385
SUMMARY
#region PAS domain
#length 848 #molecular-weight 96147 #checksum 8926

Query Match 2.4%; Score 232; DB 2; Length 848;
Best Local Similarity 27.3%; Pred. No. 3.23e-14;
Matches 42; Conservative 52; Mismatches 50; Indels 9; Gaps 7;

Db 27 PAEGIKSNPKRRHRLNLTDELRLASLLP--FPQ-DVIN-KLDKLSVLRLSVLYLRAKSF 82
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 25 PGGLTCSGKRRRQESKYIELAELISANLSDIDFNVPKDKCALLETVRQIRQIK- 83
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 83 FDVALKSSPTERRNGDNCRAANFREGNLQE--GEFLQALNGFVLVVTADALVFYASST 141
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 84 -EQG-KTISNDDDDVQKADVSSTG-QGVIDKDSLGPLQLQALDGLFVNVNREANIVFVSEN 140
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 142 IQDYLGFQSDVIHQSVYELIHTEDRAEFQRL 174
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 141 VTQYLOKQEDLVNTSVYNILHEEDRKDFLKNL 173
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 5
ENTRY
#type complete
TITLE
#formal_name Aryl hydrocarbon receptor - rat
ALTERNATE_NAMES
ORGANISM
#formal_name Rattus norvegicus #common_name Norway rat
DATE
10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change
10-Sep-1997
ACCESSIONS
S58375; S62121
REFERENCE
#authors
#journal
#title
Tissue specific expression of the rat Ah-receptor and ARNT
mRNAs.
#cross-references MUID:94344763
#accession
S58375
#molecule_type mRNA
#residues 1-853 #label CAR
#cross-references EMBL:U09000
REFERENCE
S62121
#authors
#journal
#title
Bradfield, C.A.
#cross-references MUID:94344763
#accession
S62121
#molecule_type mRNA
#residues 1-853 #label BRA
#cross-references EMBL:U09000; NID:g510268; PID:g510269
GENETICS
#gene
SUMMARY
#length 853 #molecular-weight 96220 #checksum 7010

Query Match 2.4%; Score 239; DB 2; Length 853;
Best Local Similarity 30.9%; Pred. No. 3.62e-15;
Matches 47; Conservative 46; Mismatches 51; Indels 8; Gaps 6;

Db 26 PAEGIKSNPKRRHRLNLTDELRLASLLP--FPQ-DVIN-KLDKLSVLRLSVLYLRAKSF 81
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 25 PGGLTCSGKRRRQESKYIELAELISANLSDIDFNVPKDKCALLETVRQIRQIKE 84
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 82 FDVALKSTPADRSRGDQCRQAQ--VRWDQLQEGEFLQALNGFVLVVTADALVFYASSTI 140
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 85 QKGTI-SNDDDDVQKADVSSTGCGVQDKDSL--GPLLQLQALDGLFVNVNREANIVFVSEN 141
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Db 141 QDYLGFQSDVIHQSVYELIHTEDRAEFQRL 172
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 142 VTQYLOKQEDLVNTSVYNILHEEDRKDFLKNL 173
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

GENETICS
#gene
SUMMARY
#type complete
TITLE
transcription initiation factor IID - human
ALTERNATE_NAMES
ORGANISM
#formal_name Homo sapiens #common_name man
DATE
20-Jul-1990 #sequence_revision 19-May-1995 #text_change
18-Sep-1998
ACCESSIONS
A34830; A34831; S10944
REFERENCE
#authors
#journal
#title
Functional domains and upstream activation properties of
cloned human TATA binding protein.
#cross-references MUID:90302006
#accession
A34830
#molecule_type mRNA
#residues 1-339 #label PET
#cross-references GB:M5654; NID:g339491; PID:g339492
REFERENCE
A34831
#authors
#journal
#title
Kao, C.C.; Lieberman, P.M.; Schmidt, M.C.; Zhou, Q.; Pel, R.;
Berk, A.J.
Science (1990) 248:1646-1649
```



```

#gene          sect1
CLASSIFICATION #superfamily gliadin
FEATURE
1-19          #domain signal sequence #status predicted #label SIG\
20-357        #product omega secalin #status predicted #label MAT
SUMMARY        #length 357 #molecular-weight 41439 #checksum 3741

Query Match      2.3%   Score 229;  DB 2;  Length 357;
Best Local Similarity 28.2%   Pred. No. 8,19e-14;
Matches          71;  Conservative 45;  Mismatches 117;  Indels 19;  Gaps 17;

Db    15  IITATROLNISEGLASPOQPVKEQ-S--YPOQPYPSHQPFPTPOQYSFYQPPQFPQP 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    1081 LVNGQALEPKQDAFGQEAAYVMDQKAGLYGQ-TYPAQGP-PMQGGF-HLQGSPSFNS 1137

Db    72  QQPAPIQOPFPFQOQPPFPQOQQLPLQOQPFQPOQPIPQOQOQSFQPPQQRPEQQ 131
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    1138 MX-NQMNQGNFPLQGMHPRANTMRPTNTPKOLRMQLQRL--QGCG-FLNQS-RQALE 1192

Db    132 FPQOPQGIIPQOIQOQPFPLQOQPFQPOQPSFAQOPKQIISQOQFPLQOQPFPSQPOQP 191
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    1193 LKWNPTAGGAAYVRPM-MQOQGF-LNAQMV-AQKRELLSHH-FRQQRVAMMMQOQOQ 1248

Db    192 FPQOPQGIIPQOQPSPLQOQPFPSQOQPORQOP-FPQG-PQGIIPQOQPPFPPLQPOQ 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy    1249 -QQQQQQQ-QQQQQQQQQQQQQQTQAFSPPPNVTASPSMDGLLAGTWPQAPPQOFPQPNY 1306

Db    250 PVFQOQQRPFQG 261

Qy    1307 GMGQOQDPQAFGR 1318

```

```

RESULT      11
ENTRY
TITLE       #type complete
ORGANISM    transcription factor btd - fruit fly (Drosophila sp.)
DATE        #formal_name Drosophila sp.
            18-Feb-1994 #sequence_revision 01-Dec-1995 #text_change
            24-Sep-1998
ACCESSIONS  S39356
REFERENCE   S39356
AUTHORS     Wimmer, E.A.; Jaeckle, H.; Pfeifle, C.; Cohen, S.M.
JOURNAL     Nature (1993) 366:690-694
TITLE       A Drosophila homologue of human Sp1 is a head-specific
            segmentation gene.
ACCESSION   S39356
STATUS      preliminary
MOLECULE_TYPE DNA
RESIDUES    1-644 ##label WIM
CROSS-REFS  ##cross-references EMBL:229361; NID:g441283; PID:g441284
GENETICS
GENE         FlyBase:btd
CROSS-REFS   ##cross-references FlyBase:FBgn0000233
INTRONS      245/2
SUMMARY      #length 644 #molecule-weight 68581 #checksum 6511

```

```
Query Match          2.3%   Score 230; DB 2; Length 644;
Best Local Similarity 48.3%; Pred. No. 6.01e-14;
Matches             43; Conservative 19; Mismatches 23; Indels 4; Gaps 4;
```

```
Dbb      31  QQQQHHLHMQQ-QAHHLHLSHQAQQOHMHQLTQQQQQQQQQQQQQQQQQQQQQQQQQQPQQQQ 89
           ||| : | | : | :: |:: : |::||| ||| ||| ||| ||| ||| ||| ||| |||
Qy       1211 PQQGFLNAQMWAQRSELLSHHFRCRVAMMMQQQQQQQQQQQQQQQQQQQQQQQQQ-QQTQ 1269
```

```
Dbb      90  HDFLSAAALLSAPPSSLGSGSSGSSGSSSP 118
           | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qv       1270 -AF-SPPPNVTASPSMDGLLAGPTMPOAP 1296
```

RESULT	12
ENTRY	S54522
TITLE	#type complete hypothetical protein YMR164c - yeast (Saccharomyces cerevisiae)
ALTERNATE NAMES	hypothetical protein YN8520.13c

```

ORGANISM      #formal_name Saccharomyces cerevisiae
DATE          08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
              06-Feb-1998
ACCESSIONS   S54522; S54609
REFERENCE    S54510
#authors     Hunt, S.; Bowman, S.
#submission  submitted to the EMBL Data Library, May 1995
#accesion    S54522
#molecule_type DNA
#residues_  1-758 ##label HUN
##cross-references GB:Z49705; EMBL:Z49700; NTD:g825556; PID:g825569;
               EMBL:Z49705; MIPS:YMR164C
##experimental_source strain AB972
GENETICS
#gene        SGD:MSS11
##cross-references SGD:S0004774; MIPS:YMR164c
#map_position 13R
SUMMARY      #length 758 #molecular-weight 85050 #checksum 3135

Query Match           2.3%; Score 227; DB 2: Length 758;
Best Local Similarity 44.6%; Pred.No.1.52e-13;
Matches             45; Mismatches 36; Indels 4; Gaps 3;

Db 273 QPOGSSQQTQPQHQPQHQPQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQTPY 332
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Qy 1211 QPQGFLNAQVAYQRSRELLSHHFHQRVQVMWMMQQQQQQQQQQQQQQQQQQQQQQ 1269

Db 333 PIVNPQMVFHPISNSHTG-LMPSVPTPNQFNQAQTOSM 372
: : | | : : : : : : : : | | : : | | : : | | : : | | : : | | : : | |
Qy 1270 AFSPPPVNTASPSMDGLLAGTTPMAPPP--GOFFPYOPNYGM 1308

```

```

RESULT      13
ENTRY
TITLE      A45266      #type complete
ENTRY      aryl hydrocarbon receptor - mouse
ALTERNATE_NAMES
ALTERNATE_NAMES      Ah receptor
ALTERNATE_NAMES      #formal_name Mus musculus #common_name house mouse
ALTERNATE_NAMES      21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
DATE      20-Mar-1998
ACCESSIONS
ACCESSIONS      A45266; JQ1485; A48601; A61274
REFERENCE
REFERENCE      #authors      Burbach, K.M.; Poland, A.; Bradfield, C.A.
REFERENCE      #journal      Proc. Natl. Acad. Sci. U.S.A. (1992) 89:8185-8189
REFERENCE      #title      Cloning of the Ah-receptor cDNA reveals a distinctive
REFERENCE      #note      ligand-activated transcription factor.
REFERENCE      #cross-references MUID:92390411
REFERENCE      #accession      A45266
REFERENCE      ##status      preliminary
REFERENCE      ##molecule_type mRNA: protein
REFERENCE      ##residues      1-805 ##label BUR
REFERENCE      #cross-references GB:M4523; NID:q405913; PTD:q192101
REFERENCE      #note      sequence extracted from NCBI backbone (NCBIN:112849,
REFERENCE      NCBIP:112851)
REFERENCE      JQ1485
REFERENCE      #authors      Ena, M.; Sogawa, K.; Watanabe, N.; Chujoh, Y.; Matsushita,
REFERENCE      N.; Gotoh, O.; Funae, Y.; Fujii-Kuriyama, Y.
REFERENCE      #journal      Biochem. Biophys. Res. Commun. (1992) 184:246-253
REFERENCE      #title      cDNA cloning and structure of mouse putative Ah receptor.
REFERENCE      #cross-references MUID:92231934
REFERENCE      #accession      JQ1485
REFERENCE      ##molecule_type mRNA
REFERENCE      ##residues      1-73, 'S', '75-131, 'FL', 134-170, 'HV', 173-805 ##label EMA
REFERENCE      #experimental_source HeLa-1 cells
REFERENCE      #note      residues 10-25 have been confirmed by protein sequencing
REFERENCE      A48601
REFERENCE      #authors      Schmidt, J.V.; Carver, L.A.; Bradfield, C.A.
REFERENCE      #journal      J. Biol. Chem. (1993) 268:22203-22209
REFERENCE      #title      Molecular characterization of the murine Ah gene.
REFERENCE      #note      Organization, promoter analysis, and chromosomal
REFERENCE      #note      assignment.
REFERENCE      #cross-references MUID:94012821
REFERENCE      #accession      A48601
REFERENCE      ##status      preliminary; not compared with conceptual translation

```

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 17, 1999, 19:17:22 ; Search time 1111.79 Seconds
(without alignments)
7976.786 Million cell updates/sec

Title: US-09-041-994-1
Perfect score: 4496
Sequence: 1 GCTGGATGCTGACTACAGAG.....CATTTGAGCAGGAATCTAG 4496

Scoring table: IDENTITY_NUC
2546578 seqs, 986266752 residues

Database : EST.*

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	534.4	11.9	547	34	AA488485	AA488485 ab39a08.r
2	385.2	8.6	437	26	W84822	W84822 zb53e05.r1
3	382.4	8.5	405	28	AA065270	AA065270 e08502r.T
4	370.0	8.2	425	28	AA114092	AA114092 zn66e10.r
5	353.8	7.9	427	29	AA150333	AA150333 z103e07.r
6	347.6	7.7	405	28	AA065268	AA065268 f05502m.T
7	343.4	7.6	457	40	AA920142	AA920142 vY52q12.r
8	316.4	7.0	405	28	AA065271	AA065271 f08502r.T
9	309.0	6.9	405	28	AA065272	AA065272 c08500r.T
10	300.0	6.7	464	21	T77368	T77368 yd72g08.r1
11	299.8	6.7	502	38	AA764263	AA764263 vv49f10.r
12	297.4	6.6	460	46	AI440499	AI440499 tc83e05.x
13	263.6	5.9	516	34	AA530243	AA530243 vj38c06.r
14	252.8	5.6	323	22	R28559	R28559 yH55b11.r1
15	247.8	5.5	287	27	AA045419	AA045419 zK59c01.r
16	245.0	5.4	449	26	W84775	W84775 zH53e05.s1
17	242.2	5.4	305	31	AA300819	AA300819 EST13767
18	229.2	5.1	405	28	AA065269	AA065269 d07502m.T
19	213.2	4.7	402	22	R21765	R21765 yH21a01.r1
20	193.6	4.3	377	22	R65357	R65357 yI34g11.r1
21	180.2	4.0	436	41	AI044519	AI044519 UI-R-C1-k
22	167.8	3.7	176	32	AA360136	AA360136 EST69252
23	152.8	3.4	387	27	C03704	C03704 C03704 Huma
24	151.4	3.4	209	39	AA823647	AA823647 vr69b05.s
25	147.4	3.3	487	21	T77062	T77062 yA72g08.s1
26	146.2	3.3	353	22	H04364	H04364 yJ20e11.r1
27	131.6	2.9	357	22	R21718	R21718 yH21a01.s1
28	130.2	2.9	454	22	R28355	R28355 yH55b11.s1
29	128.0	2.8	392	22	R67499	R67499 yI33g11.s1
30	128.0	2.8	392	40	AA946543	AA946543 EST202042
31	97.0	2.2	431	29	AA163058	AA163058 ms24d10.r
32	91.4	2.0	406	36	AA619146	AA619146 vc68e11.r
33	90.4	2.0	430	25	N64615	N64615 yz86e06.s1
34	90.0	2.0	368	49	AI664474	AI664474 ue63f06.r
35	87.4	1.9	524	38	AA740705	AA740705 ny98a04.s
36	83.2	1.9	385	22	R25318	R25318 yH42a03.r1
37	79.8	1.8	115	47	AI523909	AI523909 tg97h11.x
38	75.4	1.7	532	33	AA422600	AA422600 vd28e04.s
39	73.8	1.6	467	38	AA744255	AA744255 ny62g11.s
40	73.4	1.6	420	22	R66358	R66358 yI34g11.s1
41	72.8	1.6	344	39	F23060	F23060 SSC19F03 Po
42	69.2	1.5	360	50	AI678308	AI678308 tu80e01.x
43	69.0	1.5	482	29	AA187197	AA187197 zp69h04.r
44	69.0	1.5	532	40	AA939852	AA939852 vz95f01.r
45	68.4	1.5	418	42	AI077892	AI077892 oy15a04.s

ALIGNMENTS

RESULT 1
AA488485 547 bp mRNA EST 11-AUG-1997
LOCUS ab39a08.r1 Stratagene Hela cell s3 937216 Homo sapiens cDNA clone
DEFINITION IMAGE:843158 5' similar to TR:G1314285 G1314285 GRP1 ; , mRNA
sequence.
ACCESSION AA488485
NID 92215916

```
VERSION AA488485.1 GI:2215916
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 547)
Kucaba,T., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,X., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-Merck EST Project 1997
JOURNAL Unpublished (1997)
COMMENT On Jan 25, 1995 this sequence version replaced gi:637866.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m3 rev1 ET from Amersham
High quality sequence stop: 445.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Stratagene HeLa cell s3 937216"
/sex="female"
/dev_stage="Hela S3 cell line"
/lab_host="SOLR (kanamycin resistant)"
/notes="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dr. HeLa S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP XR
Vector. -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3',
adaptor sequence: 5' CTCGAGTTTTTTTTTTT 3'"
BASE COUNT 163 a 145 c 127 g 112 t
ORIGIN

Query Match 11.9%; Score 534.4; DB 34; Length 547;
Best Local Similarity 99.6%; Pred. No. 2.2e-135;
Matches 546; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

3046 TGAATCCCATGGGAATGGGGCTTAATCCCTATGGCCAGCAGCATCTAACCAACT 3105
1 TGAATCCCATGGGAATGGGGCTTAATCCCTATGGCCAGCAGCATCTAACCAACT 60

3106 GGGTTCTTGGCCGATGCATGTTCTCCATGGAACAAGTTCTCATGTCACCTCAAAATAG 3165
Db 61 GGGTTCTTGGCCGATGCATGTTCTCCATGGAACAAGTTCTCATGTCACCTCAAAATAG 120

3166 GCCTCTTTAGGAATTCCTCGATGATCTTGTGGGCCACCTTCCAACTCGGAAGGCCA 3225
Db 121 GCCTCTTTAGGAATTCCTCGATGATCTTGTGGGCCACCTTCCAACTCGGAAGGCCA 180

3226 GAGTCAGCAAGAGCATATTGGACAGCTGTCACACTTTCTCAGCAACACAGATGCAC 3285
Db 181 GAGTCAGCAAGAGCATATTGGACAGCTGTCACACTTTCTCAGCAACACAGATGCAC 240

3286 AGGCTCTGGAAGAAATGACAGAGCTTTGGGCATTCCTGAACCTGTCAATCAGGGACAGC 3345
Db 241 AGGCTCTGGAAGAAATGACAGAGCTTTGGGCATTCCTGAACCTGTCAATCAGGGACAGC 299

3346 ATTAGAGCCCAACAGGATGCTTTTCCAGAGCCCAAGCAGCAGTAATGATGGATCAGAA 3405
Db 300 ATTAGAGCCCAACAGGATGCTTTTCCAGAGCCCAAGCAGCAGTAATGATGGATCAGAA 359

3406 GGCAGGATTATATGACAGACATACCCAGCAGAGGGGCTCCATATGCAAGGAGGCTTTCA 3465
```

```
Db 360 GGCAGGATTATATGACAGACATACCCAGCAGAGGGGCTCCATATGCAAGGAGGCTTTCA 419
QY 3456 TCTTCAGGGACAATCACCATTCTTTTAACCTCTATGATGAATCAGATGAACCAAGGAGCAA 3525
Db 420 TCTTCAGGGACAATCACCATTCTTTTAACCTCTATGATGAATCAGATGAACCAAGGAGCAA 479
QY 3526 TTTTCTCTCTCAAGGAATGCACCCAGCAGCGCAACATCATGAGCCCGGACAAACACCC 3585
Db 480 TTTTCTCTCTCAAGGAATGCACCCAGCAGCGCAACATCATGAGCCCGGACAAACACCC 539
QY 3586 CAAGCAAC 3593
Db 540 CAAGCAAC 547
```

RESULT 2

```
W84822
LOCUS W84822 437 bp mRNA EST 27-JUN-1996
DEFINITION zh53e05.r1 Soares fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
Clone IMAGE:415808 5', mRNA sequence.
ACCESSION W84822
NID g1395952
VERSION W84822.1 GI:1395952
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 437)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasaki,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On May 18, 1995 this sequence version replaced gi:811263.
```

```
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 312.
Location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="GDB:1324278"
/db_xref="taxon:9606"
/clone_lib="IMAGE:415808"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DHI0B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I oligo(dT) primer [5'
AATCGAAGAAATTAATAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
```

FEATURES

Location/Qualifiers

BASE COUNT 120 a 114 c 101 g 99 t 3 others

Query Match 8.6%; Score 385.2; DB 26; Length 437;
 Best Local Similarity 96.6%; Pred. No. 1.3e-94;
 Matches 424; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

QY 3986 CCATATCAACCAAAATATGGAATGGGACAAACACAGATCCAGCCTTTGGTCGAGTGTCT 4045
 Db 1 CCATATCAACCAAAATATGGAATGGGACAAACACAGATCCAGCCTTTGGTCGAGTGTCT 60

QY 4046 AGTCCTCCCAATGAATGATGCTGCAAGATGGTCCCTCCCAAGATCCCATGATGCAA 4105
 Db 61 AGTCCTCCCAATGAATGATGCTGCAAGATGGTCCCTCCCAAGATCCCATGATGCAA 120

QY 4106 CACCGGAGGCTGATCATCTATCATCTCAGTCTCAGAAATGAAGGCTGCCATCAGGAAT 4165
 Db 121 CACCGGA-GCTGATCATCTATCATCTCAGTCTCAGAAATGAAGGCTGCCATCAGGAAT 179

QY 4166 TTGGCCAGGAACAGCTCTCTTTCCACAGCAGCTTTGCCACACAGGGGAATCCTGCAGTG 4225
 Db 180 TTGGCCAGGAACAGCTCTCTTTCCACAGCAGCTTTGCCACACAGGGGAATCCTGCAGTG 239

QY 4226 TATAGTATGTCACATGAATGGCAGCAGTGGTCACATGGGACAGATGAACATGAACCCC 4285
 Db 240 TATAGTATGTCACATGAATGGCAGCAGTGGTCACATGGGA-AAATGAACATGAACCCC 298

QY 4286 ATGCCCATGTCGGCATGCTATGGTCTGATCAGAAATACCTGCTGACATCTCTGCACC 4345
 Db 299 ATGCCCATGTCGGCATGCTATGGTCTGATCAGAAATACCTGCTGACATCTCTGNACC 358

QY 4346 AGGACCTCTTAAGG--AAACACATGACAAATGACATGACACTGACATAGGATTTGGGAAGGA 4403
 Db 359 AGGACCTCTTAAGGAAACCACTTGTNCAATGACACTGACATAGGATTTGGGAAGGG 418

QY 4404 ATCATTTGTCAGGCATCC 4422
 Db 419 ATCATTTGTCAGGCATCC 437

RESULT 3
 AA065270
 LOCUS e08502r Testis 5 Homo sapiens cDNA clone e08502 3' end, mRNA 25-SEP-1996
 DEFINITION sequence.
 ACCESSION AA065270
 NID g1929270
 VERSION AA065270.1 GI:1929270
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Guellaen, G.
 TITLE Guellaen, G. Unpublished (1996)
 JOURNAL Unpublished (1996)
 COMMENT On Sep 1, 1995 this sequence version replaced.

CONTACT: Guellaen G
 Unite INSERM 99
 INSERM
 Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
 Tel: (33)149813530
 Fax: (33)14980908
 Email: guellaen@infobiogen.fr
 This sequence derives from a clone which was selected from the cDNA
 library - Testis 5 - using a repeat of 14 CAG as probe
 Seq primer: M13 reverse.
 Location/Qualifiers
 1..405
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="e08502"
 /clone_lib="Testis 5"

FEATURES
 source

/note="Vector: pSPORT1; Site_1: MluI; Site_2: NotI; mRNA
 was prepared from human testis of a 27 years old man. cDNA
 was prepared using a 15mer oligo dt anchored by two
 degenerated bases at its 3' end and containing a NotI site
 at its 5' end. The cDNA was cloned between SalI and NotI
 sites of pSPORT1. The MluI-SalI fragment came from the
 adaptor used for the cloning. The 3' end is at the NotI
 site. cDNA corresponding to abundant species were
 eliminated from this library."

BASE COUNT 112 a 101 c 103 g 87 t 2 others
 ORIGIN

Query Match 8.5%; Score 382.4; DB 28; Length 405;
 Best Local Similarity 97.8%; Pred. No. 7.3e-94;
 Matches 397; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 3044 GGTGAATCCCATGGGAATGGGGGCTAATCCCTATGGCCCAAGCAGCAGCATCTAACCAA 3103
 Db 1 GGTGAATCCCATGGGAATGGGGGCTAATCCCTATGGCCCAAGCAGCAGCATCTAACCAA 60

QY 3104 CTGGGTTCTCTGGCCGATGGCATGTTGTCCATGGAACAAGTTTCTCATGGCACTCAAAAT 3163
 Db 61 CTGGGTTCTCTGGCCGATGGCATGTTGTCCATGGAACAAGTTTCTCATGGCACTCAAAAT 120

QY 3164 AGGCCTCTCTTAGGAATTCCTCTGGATGATCTTTGGGCCACCTTCCAACTGGGAAGC 3223
 Db 121 AGGCCTCTCTTAGGAATTCCTCTGGATGATCTTTGGGCCACCTTCCAACTGGGAAGC 179

QY 3224 CAGAGTGACAAAGAGCATTTATGGACCACTGTCACACTTCTCAGCAACACAGATGCG 3283
 Db 180 CAGAGTGACAAAGAGCATTTATGGACCACTGTCACACTTCTCAGCAACACAGATGCG 239

QY 3284 ACAGGCTCTGAAGAAATGACAGAGCTTTGGGCATCTCTGAACCTTGTCAATCAGGGACAG 3343
 Db 240 ACAGGCTCTGAAGAAATGACAGAGCTTTGGGCATCTCTGAACCTTGTCAATCAGGGACAG 299

QY 3344 GCATTAGAGCCCAACAGGATGCTTTCCAAAGGCCAAGACAGCAGCATTAATGATGATCAG 3403
 Db 300 GCATTAGAGCCCAACAGGATGCTTTCCAAAGGCCAAGACAGCAGCATTAATGATGATCAG 359

QY 3404 AAGCGAGGATTAATGACAGACATACCCAGCAGAGGGGCTCCAA 3449
 Db 360 AAGCGAGGTTTATATGGACAGACTTACCCAGCAGAGGGGCTCCCA 405

RESULT 4

AA114092

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

On Sep 1, 1995 this sequence version replaced.

Contact: Wilson RK

QY 3540 GAATGACCCACGA--GCCAATCATGATGACACCCCGGACAAACACCCCAAGCAACTTAG 3597
 Db 300 GAATGACCCACGAAGCCCAATCATGATGACACCCCGGACAAACACCCCAAGCAACTTAG 359

QY 3598 AAT-GCAGCTTCACGAGGCTGCAGGCCAGCAG 3631
 Db 360 AATGCGAGCTTCACGAGGCTGCAGGCCAGCAG 394

RESULT 6
 AA065268/c 405 bp mRNA EST 25-SEP-1996
 LOCUS f05502m Testis 5 Homo sapiens cDNA clone f05502 3' end, mRNA
 DEFINITION sequence.
 ACCESSION AA065268
 NID g1929268
 VERSION AA065268.1 GI:1929268
 WORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Guellaen,G.
 TITLE Guellaen,G. Unpublished (1996)
 JOURNAL Unpublished (1996)
 COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Guellaen G
 Unite INSERM 99
 INSERM
 Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
 Tel: (33)149813530
 Fax: (33)14980908
 Email: guellaen@infobiogen.fr
 This sequence derives from a clone which was selected from the cDNA
 library - Testis 5 - using a repeat of 14 CAG as probe
 Seq primer: -21 M13.
 FEATURES Location/Qualifiers
 source 1..405
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="f05502"
 /note="Vector: pSPOR1; Site.1: M13; Site.2: NotI; mRNA
 was prepared from human testis of a 27 years old man. cDNA
 was prepared using a 15mer oligo dT anchored by two
 degenerated bases at its 3' end and containing a NotI site
 at its 5' end. The cDNA was cloned between SalI and NotI
 sites of pSPOR1. The M13-SalI fragment come from the
 adaptor used for the cloning. The 3' end is at the NotI
 site. cDNA corresponding to abundant species were
 eliminated from this library."
 BASE COUNT 98 a 100 c 98 g 102 t 7 others
 ORIGIN

Query Match 7.7%; Score 347.6; DB 28; Length 405;
 Best Local Similarity 97.0%; Pred. No. 2.4e-84;
 Matches 350; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4131 AGTCCTCAGAAATGAAGGCTGGCCATCAGAAATTTGGCCAGGAACAGCTCTTTTCCC 4190
 Db 405 AGTCCCCANAAATGAAGGCTGGCCATCAGAAATTTGGCCAGGAACAGCTCTTTTCCC 346

QY 4191 AGCAGAGTTTGGCCACAGGGGAATCCTCAGTGTTATAGTATGGTGCATGAATGGCA 4250
 Db 345 AGCAGAGTTTGGCCACAGGGGAATCCTCAGTGTTATAGTATGGTGCATGAATGGCA 286

QY 4251 GCAGTGTGCATGGGACAGATGACATGACACCCATGCCATGCTCTGGATGCTATGG 4310
 Db 285 GCNCTGTGCATGGGACAGATGAACATGAACCCCATGCCATGCTCTGGATGCTATGG 226

QY 4311 GTCTGATCAGAATACTGTGTGACATCTCTGCACAGGACCTCTTAAGGAAACCACTGTA 4370
 Db 225 GTCTGATCAGAATACTGTGTGACATCTCTGCACAGGACCTCTTAAGGAAACCACTGTA 166

QY 4371 CAAATGACACTGCACAGGATTTGGGAAGGAATCAATTTCCAGGCATCCATCTTGA 4430
 Db 165 CAAATGACACTGCACAGGATTTGGGAAGGAATCAATTTCCAGGCATCCATCTTGA 106

QY 4431 AGAAGGACCAAGCTTTGAGCTCCATCAAGGGTATTTTAAGTGATGTCATTTGAGCAGAA 4490
 Db 105 AGAAGGACCAAGCTTTGAGCTCCATCAAGGGTATTTTAAGTGATGTCATTTGAGCAGAA 46

QY 4491 T 4491
 Db 45 T 45

RESULT 7
 AA920142 457 bp mRNA EST 20-APR-1998
 LOCUS AA920142
 DEFINITION VY52q12.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
 IMAGE:1299142 5' similar to TR:009000 O09000 P300/CBP/CO-INTEGRATOR
 PROTEIN: ; mRNA sequence.
 ACCESSION AA920142
 NID 93066921
 VERSION AA920142.1 GI:3066921
 WORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 457)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1407511.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:680190
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 410.
 FEATURES Location/Qualifiers
 source 1..457
 /organism="Mus musculus"
 /strain="C57BL/6 x CBA"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:1299142"
 /clone_lib="Stratagene mouse lung 937302"
 /sex="female"
 /tissue_type="lung"
 /dev_stage="6-8 month old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: lung; Vector: pBluescript SK-; Site.1:
 ECORI; Site.2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. 6-8 month old female lung and 1.5 year old male
 lung were source of mRNA. Average insert size: 1.5 kb;
 Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGCAGG
 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 120 a 119 c 122 g 96 t

ORIGIN

	Query Match	7.6%	Score 343..4;	DB 40;	Length 457;
	Best Local Similarity	84.5%;	Pred. No. 3.6e-83;		
	Matches 386;	Conservative 0;	Mismatches 71;	Indels 0;	Gaps 0;
QY	3058	GGGAATGGGGGCTTAATCCCTATGGCCAAAGCAGCAGCATCTAACCAACTGGTTCCTGGCC	3117		
DB	1	GGGAATGGGAGTCAATCCCTATAGCCAGCAGAGTCCGCTCTAACCAACAGGTTCTCTGGCC	60		
QY	3118	CGATGGCATCTTGCCATGGAAACAAGTTTCTCATGGCAGCTCAAAATAGGCTCTCTCTTAG	3177		
DB	61	AGAGGGCATGCTCTCTATGGAACAAGGCTCTCAGGGCTCAAAATAGGCTCTCTCTTAG	120		
QY	3178	GAATTCCTCGGATGATCTTGTGGGCCACCTTCCAACTGGAGGCCAGAGTACGAAAG	3237		
DB	121	AAACTCTCTGGATGATCGTGTGGGCCACCTTCTAACCGCAGAGGCCAGAGTACGAGAG	180		
QY	3238	AGCATTATGGACAGCGTGCCACACTCTTCTCAGCAACACAGATGCGACAGCCCTGGAAGA	3297		
DB	181	AGCTCTGCTGGACAGGTGCACACACTCTTGAGCAACACAGATGCCACAGCTCTGGAGGA	240		
QY	3298	AATTGACAGAGCTTTGGGCATTCTGAACTGTCAATCAGGCAGCAGCATATAGAGCCCAA	3357		
DB	241	GATCGACAGGGCCCTTGGGAAATCTTGAGCTCGTGAATCAGGGACAAGCTTTGGAGTCCAA	300		
QY	3358	ACAGGATGCTTTCCAGGCCCAAGACAGCAGTAATCATGATCAGAACAGGCAGGATTATA	3417		
DB	301	ACAGGATGTTTTCCAGGCCCAAGACAGCAGTAATCATGATCAGAACAGGCCTGCACATATA	360		
QY	3418	TGGACACAGACATATCCACACAGAGGGCCCTCCAATGCAAGAGGAGGCTTTCATCTTCAGGGACA	3477		
DB	361	TGCACAGACATATCCAGCTCAGGGTCTCTCCCTCTCAGGAGGCTTTAACTCTCAGGACA	420		
QY	3478	ATCACCATCTTTTAACCTCTATGATGAATCAGATGAAC	3514		
DB	421	GTCAACATCTTTTAACCTCTATGATGGCTCAGATTAG	457		

RESULT	8
AA065271	405 bp mRNA EST 25-SEP-1996
LOCUS	f08502r Testis 5 Homo sapiens cDNA clone f08502 3' end, mRNA
DEFINITION	sequence.
ACCESSION	AA065271
NID	g1929271
SOURCE	AA065271.1 GI:1929271
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 405) Guellaen, G. Unpublished (1996)
REFERENCE	Unpublished (1996)
AUTHORS	On Sep 1, 1995 this sequence version replaced.
TITLE	
JOURNAL	
COMMENT	

```

/clone="f08502"
/note="Vector: pSPORT1; Site_1: MluI; Site_2: NotI; mRNA
was prepared from human testis of a 27 years old man. cDNA
was prepared using a 15mer oligo dT anchored by two
degenerated bases at its 3'-end and containing a NotI site
at its 5'-end. The cDNA was cloned between SalI and NotI
sites of pSPORT1. The MluI-SalI fragment come from the
adaptor used for the cloning. The 3' end is at the NotI
site. cDNA corresponding to abundant species were
eliminated from this library."
BASE COUNT      117 a   124 c    91 g       72 t        1 others
ORIGIN

Query Match          7.0%; Score 316.4; DB 28; Length 405;
Best Local Similarity 91.8%; Pred. No. 8.2e-76;
Matches 368; Conservative 0; Mismatches 27; Indels 6; Gaps 3;

QY 3796 GAGGCTGGCTATGATGTCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCA 3855
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3 GATGTCAGCAGCAGCAGCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAAGCA 62

QY 3856 ACAGCAACAGCAACAGCAGCAACAGCAGCAAAACCAGGCCTTCAGCCCACTCTAATGT 3915
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 ACAGCAACAGCAACAGCAGCAACAGCAGCAAAACCAGGCCTTCAGCCCACTCTAATGT 122

QY 3916 GACTGCTCCCCCAGCATGGATGGGCTTTTGGCAGAGCCACCAATGCCACAAGCTCCTCC 3975
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 123 GACTGCTCCCCCAGCATGGATGGGCTTTTGGCAGAGCCACCAATGCCACAAGCTCCTCC 182

QY 3976 GCAACAGTTTTCCATATCAACCAAATTATGGAATGGGACAAACACAGATCCAGCCTTTGG 4035
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 GCAACAGTTTTCCATATCAACCAAATTATGGAATGGGACAAACACAGATCCAGCCTTTGG 242

QY 4036 TCGAGTGTTAGTCTCTCCCNAATGCAATGATGTCGTCAAGAATGGTGCTCCCGAGAATCC 4095
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 TCGAGTGTTAGTCTCTCCCNAATGCAATGATGTCGTCAAGAATGGTGCTCCCGAGAATCC 302

QY 4096 CATGATGCAACACCCGAGCTGCATCCATC---TATCAGTCTCAGAAATGAAGGCT-- 4152
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 303 CATGATGCAACACCCGAGCTGCATCCATCTTTATTCAGTTCTCAGAAATGAAGGGTG 362

QY 4152 GGCCATCAGGAATTTGG--CCAGGAACAGCTCCTTTTCC 4190
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 GGCCATCAGGAATTTGGGCCAGGGAACAGCTCCTTTCC 403

RESULT 9
LOCUS AA065272
DEFINITION c08500r Testis 5 Homo sapiens cDNA clone c08500 3' end, mRNA
sequence.
ACCESSION AA065272
NID 91929272
VERSION AA065272.1 GI:1929272
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 405)
TITLE Guellaen,G. Unpublished (1996)
JOURNAL Guellaen,G. Unpublished (1996)
COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Guellaen G
Unite INSERM 99
INSERM
Unite INSERM 99, Hopital Henri Mondor, 94010 Creteil, France
Tel: (33)149813530
Fax: (33)14980908
```


Db	133	GTCTCTCCCAAGTGAATGATGTCTATCAAGAATGGGGCTTCCCAAGAATGCCATGGTGACG	192
QY	4107	ACCCGAGAGCTGCATCCATCTATTCAGTCTCTCAGAAATGAAGGCTTGSCCATCAGGAATT	4166
Db	193	ATCCTCAGCCACACCCATGTATTCAGCTTCAGATATGAAGGGTGCCCTCAGNGAAC	252
QY	4167	TGCCCAGGAACAGCTCTCTTTCCACGACGAGTGTGCCCCACAGGGGAATCTCTGCAGT-G	4225
Db	253	TGCCCAGGAATGCTCTCTTCCCCCAGCAGCAGTTTGCTCTCCCAGGGGAACCCCTGCAGTCC	312
QY	4226	TATAGTATGGTGACATAGAAATGGCAGCAGTGGTTCACATGGGACAGATGAACATGAACCCC	4285
Db	313	TACAACATGGTGCATATGAACAGCAGCGGTGGGCACCTTGGGACAGATGGGCATGACCCCC	372
QY	4286	ATGCCCATGTCCTGGCATGCTATGGGTCTGTGATCAGAAATACTGCTGACATCTCTCTGCACC	4345
Db	373	ATGCCCATGTCCTGGCATGCCATGGCCACCGATCAGAAATACTGCTGACATCTCTCCCTAGT	432
QY	4346	AGGACCTCTTAAGGAAACCACTGTACAAATGA	4377
Db	433	GAGACTGACTGTACAGATGACACTGGACACAGA	464

LOCUS	RESULT 12	AI440499/c	LOCUS	AI440499	460 bp	mrna	EST	18-MAR-1999
433	GAGACAGCAGTGTGACACAGGACAGGA	404	tc83e05.x1	NCI_CGAP_CLL1	Homo sapiens	CDNA clone	IMAGE:207268	3'
						similar to	TR:015406	CAGH16.1, mRNA sequence.

Species: Homo sapiens
 Organism: Homo sapiens
 Taxonomy: Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eukaryota; Primates; Catarrhini; Hominidae; Homo.
 Reference: 1 (bases 1 to 460)
 Authors: NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Journal: Unpublished (1997)
 Comment: On Jun 5, 1998 this sequence version replaced gi:3188300.

```

Insert Length: 1425      Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 449.
Location/Qualifiers
1. 460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2072768"
/clone_lib="NCI-CGAP-CLL1"
/tissue_type="B-cell, chronic lymphocytic leukemia"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TCTTACCAATCGAAGTGGGAGCGCCGATGCTTTTTTTTTTTTTTTTTT
t 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into

```

the Not I and Eco RI sites of the modified pMT3 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo.

BASE COUNT
ORIGIN

129 a 107 c 100 g 124 t

Query Match 5.6%; Score 297.4; DB 46; Length 460;
Best Local Similarity 99.7%; Pred. No. 1.4e-70;
Matches 298; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4193 CAGCAGTTGGCCACAGGGAATCCTGCAGTGTATAGTATGTTGCACATCAATGGCAGC 4252

Db 460 CAGCAGTTGGCCACAGGGAATCCTGCAGTGTATAGTATGTTGCACATCAATGGCAGC 401

QY 4253 AGTGGTCACATGGGACAGATGAACATGAACCCCATGTCCTGGCATGCCCTATGGGT 4312

Db 400 AGTGGTCACATGGGACAGATGAACATGAACCCCATGTCCTGGCATGCCCTATGGGT 341

4313 CCTGATCAGAATACTGCTGCATCTCTGCACAGGACCTCTTAAGGAACCACTGTACA 4372

Db 340 CCTGATCAGAATACTGCTGCATCTCTGCACAGGACCTCTTAAGGAACCACTGTACA 281

QY 4373 AATGACATGCTAGGATTTATGGAGGAATCATTTGTCAGGCATCCATCTTGGAG 4432

Db 280 AATGACATGCTAGGATTTATGGAGGAATCATTTGTCAGGCATCCATCTTGGAG 221

QY 4433 AAAGGACCATGTTGAGCTCCATCAAGGGTATTTAGTGTATGTCATTGACGAGGAAT 4491

Db 220 AAAGGACCATGTTGAGCTCCATCAAGGGTATTTAGTGTATGTCATTGACGAGGAAT 162

RESULT 13
AA530243

LOCUS AA530243 516 bp mRNA EST 22-JUL-1997
DEFINITION vj38c06.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
clone IMAGE:931306 5' similar to TR:G1336160 G1336160 STERIOD
RECEPTOR COACTIVATOR 1A.; mRNA sequence.

ACCESSION
AA530243

NID 92272949

VERSION AA530243.1 GI:2272949

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project

JOURNAL

COMMENT Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1393273.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:536226

Seq primer: -28m13 rev1 ET from Amersham

High quality sequence stop: 335.

FEATURES
source

1. .516
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:931306"

/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"

/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site: 1:
EcoRI; Site: 2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dT. Average
insert size: 1.5 kb. Uni-ZAP XR vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT 121 a 150 c 141 g 104 t
ORIGIN

Query Match 5.9%; Score 263.6; DB 34; Length 516;
Best Local Similarity 80.6%; Pred. No. 2.6e-61;
Matches 308; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 3996 CAAATTATGTAATGGGACACACACACATCGAGCCTTTGGTCGAGTGTCTAGTCTCTCCA 4055

Db 1 CAAATTATGTAATGGGACACACACACATCGAGCCTTTGGTCGAGTGTCTAGTCTCTCCA 60

QY 4056 ATGCAATGATGCTCTCAAGATGGTCCCTCCCAAGATCCCATGATCAACACCCGAGG 4115

Db 61 GTGCAATGATGCTCTCAAGATGGTCCCTCCCAAGATCCCATGATCAACATCTCTCAGC 120

QY 4116 CTGCAATCTCTATCAGTCTCTCAGAAATGAAGGCTGGCCATCAGAAATTTGGCCAGGA 4175

Db 121 CCACACCCATGATCAGCCTTCAGATATGAAGGGTGGCGTCAGGACACCTGGCCAGGA 180

QY 4176 ACAGCTCCTTTTCCAGCAGCAGTTTCCCAACAGGGAATCTTCAGTGTATATAGTATGG 4235

Db 181 ATGGCTCTTCCCGCAGCAGCAGTTTCTCCCGAGGGAACCTTCGACGCTTACAACATGG 240

QY 4236 TGCACATGAATGGCAGCAGTGGTCACATGGGACAGATGAACATCAACCCATGCCATGT 4295

Db 241 TGCATATGAACAGCAGCGGTGGGCACCTTGGGACAGATGGCCATGATACCCCATGCCATGT 300

QY 4296 CTGGCATGCCCTATGGTCTCTGATCAGAAATACTCTGTCACATCTCTGCACACGAGACCTCTT 4355

Db 301 CTGGCATGCCCTATGGGCGCCGATCAGAAATACTCTGTCACATCTCTGTCAGTGGGACTGACT 360

QY 4356 AAGGAACACCTGTACAAATGA 4377

Db 361 GTACAGATGACACTGCACAGGA 382

RESULT 14

R28559

LOCUS

DEFINITION

R28559 323 bp mRNA EST 25-APR-1995

YH5511.r1 Soares placenta Nb2HP Homo sapiens cDNA clone

IMAGE:133629 5' similar to contains TAR1 repetitive element ;, mRNA

sequence.

ACCESSION

R28559

NID 9784694

VERSION

R28559.1 GI:784694

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 323)

AUTHORS

Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,

Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,

Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,

Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and

Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 630
High quality sequence stops: 314

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 630 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 314.

Location/Qualifiers

FEATURES

source

1. .323
/organism="Homo sapiens"
/db_xref="GDB:539364"
/db_xref="taxon:9606"
/clone="IMAGE:133629"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
AAGTGAAGAATTCGCGCGCAGGAATTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo. "

BASE COUNT
ORIGIN

86 a 73 c 86 g 73 t 5 others

Query Match 5.6%; Score 252.8; DB 22; Length 323;

Best Local Similarity 97.5%; Pred. No. 1.8e-58;

Matches 277; Conservative 0; Mismatches 4; Indels 3; Gaps 2;

QY 4137 CAGAAATGAAGGCTGCCATCAGGAAATTTGGCCAGGACAGCTCTTTTCCACGAGC 4196

Db 1 CAGAAATGAAGGCTGCCATCAGGAAATTTGGCCAGGACAGCTCTTTTCCACGAGC 60

QY 4197 AGTTGCCACACGAGGGAATCTCTGAGTGTATAGTGTGCACATGAATGGCAGCAGTG 4256

Db 61 AGTTGCCACACGAGGGAATCTCTGAGTGTATAGTGTGCACATGAATGGCAGCAGTG 120

QY 4257 GTCACATGGGACAGATGACATGAACCCATGCCCATGTCTGGCATGCCTATGGGTCTTG 4316

Db 121 GTCACATGGGACAGATGACATGAACCCATGCCCATGTCTGGCATGCCTATGGGTCTTG 180

QY 4317 ATCAGAAATACTCTGACATCTCTGACAGGACCTCTTAAGGAACACCACTGTACAAA-T 4375

Db 181 ATCAGAAATACTCTGACATCTCTGACAGGACCTCTTAAGGAACACCACTGTACAAAAT 240

QY 4376 GACATGCTACTAGGATATTGGG--AAGGAATCATGTTGCCAGG 4417

Db 241 GACATGCTACTAGGATATTGGGAGGAGGANTCATTTGTTCCAGG 284

RESULT 15

AA045419

LOCUS

DEFINITION

2k59c01.r1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone

IMAGE:487104 5', mRNA sequence.

AA045419

ACCESSION

AA045419

NID

VERSION

AA045419.1 GI:1523621

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 {bases 1 to 287}

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoso, S., Dietrich, N., DuBuque, T., Favellio, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevasakis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.

TITLE

Generation and analysis of 280,000 human expressed sequence tags

JOURNAL

Genome Res. 6 (9), 807-828 (1996)

MEDLINE

97044478

COMMENT

On Jan 25, 1995 this sequence version replaced gi:637753.

Contact: Wilton RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1867 Std Error: 0.00

Seq primer: -28M13 rev2 from Anersham

High quality sequence stop: 232.

Location/Qualifiers

1. .287

/organism="Homo sapiens"

/db_xref="GDB:3760818"

/db_xref="taxon:9606"

/clone="IMAGE:487104"

/clone_lib="Soares_pregnant_uterus_NbHPU"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;

Site_2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5',

AAGTGAAGAATTCGCGCGCAGGAATTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

BASE COUNT 78 a 77 c 68 g 63 t 1 others

ORIGIN

Query Match 5.5%; Score 247.8; DB 27; Length 287;

Best Local Similarity 97.9%; Pred. No. 4e-57;

Matches 282; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 3985 TCCATATCAACCAATTTATGGAAT-GGGACAACACACAGATCCAGCCTTTGTCAGTGT 4043

Db 1 TCCATATCAACCAATTTATGGAATCGGGACAACAACAGATCCAGCCTTTGTCAGTGT 60

QY 4044 CTAGTCTCTCCCAATGCAATGATGTCGTCAGGAATGGTCCCTCCAGAAATCCCATGATGC 4103

Db 61 CTAGTCTCTCCCAATGCAATGATGTCGTCAGGAATGGTCCCTCCAGAAATCCCATGATGC 120

QY 4104 AACACCCGAGGCTGCATCCATCTATCATGTCCTCAGAAATGAAGGGCTGGCCATCAGGAA 4163

Db 121 AACACCCGAGGCTGCATCCATCTATCATGTCCTCAGAAATGAAGGGCTGG-CATCAGGAA 179

QY 4164 ATTTGGCCAGGACAGCTCTTTTCCAGCAGCAGTGTGCCACACAGGGGAATCTGCAG 4223

Db 180 ATTTGGCCAGGACAGCTCTTTTCCAGCAGCAGTGTGCCACACAGGGGAATCTGCAG 239

QY 4224 TGTATAGTATGTCACATGAA-TGCGACAGTGTGTCACATGGGACAG 4270

Db 240 TGTATAGTATGTCACATGAAATTTGGCAGCAGTGTGTCACATGGGACAG 287

Search completed: September 18, 1999, 00:52:22

Job time: 20100 sec

This Page Blank (uspto)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

This Page Blank (uspto)